



Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Engenharia Eletrônica

# **Análise comparativa de algoritmos de aprendizado profundo para o auxílio no diagnóstico de esquizofrenia a partir da classificação de MRI cerebral**

Autores: Thiago Gomes de Sousa Bezerra  
Gabriel Santos Silva Araújo

Orientador: Renan Utida Barbosa Ferreira, Dr

Brasília, DF  
2021





Thiago Gomes de Sousa Bezerra  
Gabriel Santos Silva Araújo

**Análise comparativa de algoritmos de aprendizado  
profundo para o auxílio no diagnóstico de esquizofrenia a  
partir da classificação de MRI cerebral**

Monografia submetida ao curso de graduação  
em (Engenharia Eletrônica) da Universidade  
de Brasília, como requisito parcial para ob-  
tenção do Título de Bacharel em (Engenharia  
Eletrônica).

Universidade de Brasília - UnB

Faculdade UnB Gama - FGA

Orientador: Renan Utida Barbosa Ferreira, Dr

Brasília, DF

2021

---

Thiago Gomes de Sousa Bezerra

Gabriel Santos Silva Araújo

Análise comparativa de algoritmos de aprendizado profundo para o auxílio no diagnóstico de esquizofrenia a partir da classificação de MRI cerebral/ Thiago Gomes de Sousa Bezerra

Gabriel Santos Silva Araújo. – Brasília, DF, 2021-

81 p. : il. (algumas color.) ; 30 cm.

Orientador: Renan Utida Barbosa Ferreira, Dr

Trabalho de Conclusão de Curso – Universidade de Brasília - UnB

Faculdade UnB Gama - FGA , 2021.

1. Esquizofrenia. 2. MRI. I. Renan Utida Barbosa Ferreira, Dr. II. Universidade de Brasília. III. Faculdade UnB Gama. IV. Análise comparativa de algoritmos de aprendizado profundo para o auxílio no diagnóstico de esquizofrenia a partir da classificação de MRI cerebral

CDU 02:141:005.6

---

Thiago Gomes de Sousa Bezerra  
Gabriel Santos Silva Araújo

# **Análise comparativa de algoritmos de aprendizado profundo para o auxílio no diagnóstico de esquizofrenia a partir da classificação de MRI cerebral**

Monografia submetida ao curso de graduação em (Engenharia Eletrônica) da Universidade de Brasília, como requisito parcial para obtenção do Título de Bacharel em (Engenharia Eletrônica).

Trabalho aprovado. Brasília, DF, 28 de Dezembro de 2021:

---

**Renan Utida Barbosa Ferreira, Dr**  
Orientador

---

**Cristiano Jacques Miosso, Dr**  
Convidado 1

---

**Marcus Vinícius Chaffim Costa, Dr**  
Convidado 2

Brasília, DF  
2021



### **Dedicatória.**

*Este trabalho é dedicado à todos que nos incentivaram e motivaram durante nossa jornada acadêmica: Nossos pais, irmãos, professores e amigos que tanto nos apoiam.*





# Agradecimentos

Primeiramente a Deus e aos nossos familiares que nos auxiliaram e sempre estiveram ao nosso lado nos obstáculos ao longo do curso. Gostaria de agradecer também aos professores, em especial ao orientador Renan Utida Barbosa Ferreira pela colaboração, empenho e paciência no desenvolvimento deste trabalho. Aos amigos pelo apoio e incentivo nos momentos difíceis e de ausência.



# Resumo

A esquizofrenia é uma doença psiquiátrica crônica de difícil compreensão que perturba e desabilita cerca de 1% da população mundial, o que pode tornar o paciente dependente da ajuda de terceiros para tarefas cotidianas devido aos seus sintomas, caso não ocorra o devido tratamento. Em muitos casos estes sintomas não evidenciam o diagnóstico da doença, o que a torna difícil. Uma alternativa é a análise morfológica do cérebro, através de diferenças perceptíveis nas regiões cerebrais de pacientes portadores de esquizofrenia comparados a indivíduos saudáveis. Neste contexto, busca-se analisar métodos de classificação automática, com arquitetura de redes neurais convolucionais, para distinguir pacientes portadores de esquizofrenia de pacientes saudáveis por meio de imagens de ressonância magnética.

Na literatura, existem diversas arquiteturas de redes neurais disponíveis para estudo. Porém, por apresentarem aplicações com imagens de ressonância magnética, além de bons resultados e com a maior disponibilidade de informações para a reprodutibilidade, como as etapas de pré-processamento e parâmetros de treinamento das redes, as redes exploradas foram as arquiteturas FayNet e ResNet, com as imagens disponibilizadas pelo banco de dados virtual *SchizConnect*.

Em uma primeira etapa foi realizado o pré-processamento das imagens com o *software* Clínica, aplicando-se a técnica de processamento de dados *t1-linear*, que realiza uma normalização MNI para padronização das imagens para posteriormente utilizar a biblioteca *Nilearn* e então obter os cortes das imagens. Os cortes 15 e 20 foram selecionados das imagens tridimensionais normalizadas.

As redes foram implementadas e validadas por métricas de desempenho e *k-fold* (com  $k = 7$ ), obtendo uma acurácia de  $80,0\% \pm 2,29\%$  e *F2-score* de  $80,0\% \pm 2,40\%$ , para o modelo FayNet, acurácia de  $85,2\% \pm 2,47\%$  e *F2-score* de  $85,4\% \pm 0,03\%$  na arquitetura criada baseado na ResNet18. Com o intuito de buscar melhorias na acurácia de classificação, a FayNet foi modificada, objetivando-se aumentar o desempenho da rede, dando origem assim a Faynet 2.0 que obteve uma acurácia de  $88,7\% \pm 4,07\%$  e *F2-score* de  $85,7 \pm 3,40\%$ .

**Palavras-chaves:** Esquizofrenia, redes neurais, diagnóstico, MRI.



# Abstract

Schizophrenia is a chronic psychiatric disease that is difficult to understand and that disturbs and disables about 1% of the world population, which can make the patient dependent on the help of others for daily tasks due to their symptoms, if the proper treatment is not followed. . In many cases, these symptoms do not show the diagnosis of the disease, which makes it difficult. An alternative is the morphological analysis of the brain, through noticeable differences in the brain regions of patients with schizophrenia compared to healthy individuals. In this context, we seek to analyze automatic classification methods, with convolutional neural network architecture, to distinguish patients with schizophrenia from healthy patients through magnetic resonance images.

In the literature, there are several architectures of neural networks available for study, however, for presenting applications with magnetic resonance images, in addition to good results with greater availability of information for reproducibility, such as the pre-processing steps and training parameters of the networks . The networks explored were the FayNet and ResNet architectures, with the images made available by the virtual database *Schiz-Connect*.

In a first step, the images were pre-processed with the *software* Clínica, applying the *pipeline t1-linear*, which performs an MNI normalization for standardizing the images and later using a *Nilearn* library and then get the image slices. Cuts 15 and 20 were selected from normalized three-dimensional images.

The networks were implemented and validated by performance metrics and *k-fold* (with  $k = 7$ ), obtaining an accuracy of  $80.0\% \pm 2.29\%$  and F2-*score* of  $80.0\% \pm 2.40\%$  , for the FayNet model, accuracy of  $85.2\% \pm 2.47\%$  and F2- *score* of  $85.4\% \pm 0.03\%$  in the architecture created based on ResNet18. In order to improve the accuracy of classification, FayNet was modified, aiming to increase the performance of the network, thus giving rise to Faynet 2.0 which obtained an accuracy of  $88.7\% \pm 4.07\%$  and F2-*score* of  $85.7 \pm 3.40\%$ .

**Key-words:** Schizophrenia, neural networks, diagnosis, Magnetic resonance imaging.



# Lista de ilustrações

Figura 1 – Exemplo de uma operação de convolução de uma imagem 6x6 por um filtro 2x2. Fonte: Autor . . . . .	31
Figura 2 – Modelo de um <i>perceptron</i> . Fonte: (ZHANG et al., 2020) . . . . .	33
Figura 3 – Arquitetura base de uma CNN. Fonte: Adaptado de (VERGARA, 2018). . . . .	35
Figura 4 – Comparação entre as funções de perda utilizando ReLU (linha sólida) e tanh (linha tracejada) realizada por Krizhevsky em um problema de classificação. Demonstrando a maior velocidade no treinamento utilizando a RELU. Fonte: (KRIZHEVSKY; SUTSKEVER; HINTON, 2017a) . . . . .	37
Figura 5 – Filtro 2x2 deslocando-se na imagem com passo de 2, de forma que entre os 4 elementos da janela é escolhido o de maior valor. Fonte: Adaptado de (KARPATHY, 2018). . . . .	38
Figura 6 – A esquerda: Uma rede neural padrão com 2 camadas ocultas. À direita: um exemplo de rede reduzida produzida pelo <i>dropout</i> . Unidades cruzadas foram descartadas. Fonte: (SRIVASTAVA et al., 2014) . . . . .	39
Figura 7 – Arquitetura da rede LeNet. Fonte: (Lecun et al., 1998). . . . .	40
Figura 8 – Arquitetura da rede Alex Net. Fonte: (KRIZHEVSKY; SUTSKEVER; HINTON, 2012). . . . .	41
Figura 9 – Arquitetura da rede ZFNet. Fonte: (DAS, 2017). . . . .	42
Figura 10 – Arquitetura da rede VGGNet. Fonte: (DHOMNE; KUMAR; BHAN, 2018). . . . .	43
Figura 11 – Bloco de construção ResNet. Fonte: Traduzido de (Budhiman; Suyanto; Arifianto, 2019). . . . .	43
Figura 12 – Arquitetura da FayNet. Fonte: (VERGARA, 2018) . . . . .	44
Figura 13 – Exemplo de partição na validação cruzada com k= 5. Fonte: Autor . . . . .	46
Figura 14 – Exemplo da separação de <i>folds</i> para treinamento e validação de um modelo na validação cruzada com k = 5. Fonte: Autor . . . . .	47
Figura 15 – Exemplo ROC e área sobre a curva AUC próxima à ideal. Fonte: Autor . . . . .	49
Figura 16 – Organização dos dados no SchizConnect. Fonte: Traduzido de (WANG et al., 2018) . . . . .	52
Figura 17 – Visualização 3D a partir do software 3D <i>SOCR</i> de uma imagem de ressonância magnética do banco de dados do <i>SchizConnect</i> . Fonte: Autor. . . . .	53
Figura 18 – Detalhes do modelo Faynet vs Faynet2.0. Fonte: Autor. . . . .	57
Figura 19 – Diagrama contendo modelo da ResNet18 com os <i>kernels</i> utilizados para convolução e <i>pooling</i> . Fonte: Autor. . . . .	58
Figura 20 – Etapa de pré-processamento da imagem 1. Fonte: Autor. . . . .	61
Figura 21 – Etapa de pré-processamento da imagem 2. Fonte: Autor. . . . .	62

Figura 22 – Gráficos de desempenho da FayNet com o banco de dados FBIRN. Fonte: Autor. . . . .	64
Figura 23 – Gráficos de desempenho da FayNet 2.0 com o banco de dados FBIRN. Fonte: Autor. . . . .	65
Figura 24 – Curva ROC dos modelos com o banco de dados FBIRN. Fonte: Autor.	65
Figura 25 – Resultados do treinamento e validação da FayNet utilizando imagens disponibilizadas pelo SchizConnect. Fonte: Autor. . . . .	67
Figura 26 – Resultados do treinamento e validação FayNet 2.0 utilizando imagens disponibilizadas pelo SchizConnect. Fonte: Autor. . . . .	67
Figura 27 – Curva ROC dos modelos com o banco de dados SchizConnect. Fonte: Autor. . . . .	68
Figura 28 – Resultados do treinamento e validação ResNet18 utilizando imagens disponibilizadas pelo SchizConnect. Fonte: Autor. . . . .	70
Figura 29 – ROC e AUC referente ao teste do modelo da ResNet18. Fonte: Autor. .	70



# Lista de tabelas

Tabela 1 – Matriz de confusão . . . . .	47
Tabela 2 – Organização detalhada dos dados no SchizConnect . Fonte: Adaptado de (WANG et al., 2018) . . . . .	52
Tabela 3 – Arquitetura da rede FayNet com os parâmetros para o banco de dados FBIRN. . . . .	63
Tabela 4 – Arquitetura da rede FayNet 2.0 com os parâmetros para o banco de dados FBIRN. . . . .	63
Tabela 5 – Comparação das métricas de desempenho da FayNet vs FayNet2.0 com imagens FBIRN. . . . .	64
Tabela 6 – Arquitetura da rede FayNet com os parâmetros para o banco de dados SchizConnect. . . . .	66
Tabela 7 – Arquitetura da rede FayNet com os parâmetros para o banco de dados SchizConnect. . . . .	66
Tabela 8 – FayNet original vs FayNet2.0 utilizando imagens disponibilizadas pelo SchizConnect. Fonte: Autor. . . . .	68
Tabela 9 – Métricas de desempenho para avaliação do modelo treinado da ResNet18. . . . .	69
Tabela 10 – Tabela de comparação dos modelos descritos. Fonte: Autor. . . . .	71



# Lista de abreviaturas e siglas

AUC	Area Under the Curve
BCE	Binary cross
BIDS	Brain Imaging Data Structure
CE	Cross entropy
CNN	Convolutional Neural Network
CT	Computed Tomography
DSM	Diagnostic and Statistical Manual of Mental Disorders
DSM - 5	Statistical Manual of Mental Disorders, fifth edition
ESQ	Esquizoafetivo
ESQA	Esquizofrenia ampla
ESQR	Esquizofrenia rigorosa
FBIRN	Function Biomedical Informatics Research Network
FN	Falso negativo
FP	Falso positivo
GPU	Graphic processing unit
IEA	Irmão da esquizofrenia ampla
IEE	Irmão da esquizofrenia estrita
INTC	Irmão de nenhum transtorno conhecido
KNN	K-Nearest Neighbors
LSTM	Memória de Longo-Curto Prazo
MLP	Multilayered perceptron
MRI	Magnetic Resonance Imaging
NIFTI	Neuroimaging Informatics Technology Initiative

NTC	Nenhum distúrbio conhecido
ReLU	Rectified linear unit
RNA	Redes Neurais Artificiais
RNN	Rede Neural Recorrente
ROI	Region Of Interest
ROC	Receiver Operating Characteristic Curve
SGD	Stochastic gradient descent
SPM	Statistical Parametric Mapping
SVM	Support Vector Machine
TB	Transtorno Bipolar
TFP	Taxa de falsos positivos
TN	Negativo verdadeiro
TP	Positivo verdadeiro
TVP	Taxa de verdadeiros positivos

# Lista de símbolos

$\beta$	Constante de ponderação da métrica $F\beta$ -score
$\Sigma$	Somatório
$\sigma$	Função de ativação sigmóide.



# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>23</b>
<b>1.1</b>	<b>Contextualização</b>	<b>23</b>
<b>1.2</b>	<b>Definição do problema</b>	<b>23</b>
<b>1.3</b>	<b>Justificativa</b>	<b>24</b>
<b>1.4</b>	<b>Objetivos</b>	<b>25</b>
1.4.1	Objetivos Gerais	25
1.4.2	Objetivos Específicos	25
<b>1.5</b>	<b>Estrutura da Dissertação</b>	<b>25</b>
<b>2</b>	<b>REFERENCIAL TEÓRICO</b>	<b>27</b>
<b>2.1</b>	<b>Esquizofrenia</b>	<b>27</b>
2.1.1	Diagnóstico	27
2.1.2	Neuro Imageamento	28
2.1.2.1	Imagens de ressonância Magnética - MRI	28
2.1.3	Alterações cerebrais observadas em pacientes portadores de esquizofrenia	29
<b>2.2</b>	<b>Processamento digital aplicado a imagens</b>	<b>29</b>
2.2.1	Processamento digital de imagens	30
2.2.2	Convolução de imagens digitais	30
<b>2.3</b>	<b>Aprendizado de Máquinas</b>	<b>31</b>
2.3.1	Classificação automática aplicada a imagens	31
2.3.2	Redes Neurais artificiais - RNA	32
2.3.2.1	Redes de camada única	33
2.3.2.2	<i>Deep learning</i> e Redes de múltiplas camadas	33
2.3.2.3	Redes Neurais convolucionais - CNN	34
2.3.2.3.1	Camada convolucional	35
2.3.2.3.2	Camada de Não-linearidades	36
2.3.2.3.3	Normalização	36
2.3.2.3.4	Camada de Pooling	37
2.3.2.3.5	Camada Totalmente Conectada	38
2.3.2.3.6	Camada de <i>Dropout</i>	39
2.3.2.4	Tipos de CNN's	39
2.3.2.4.1	LeNet	40
2.3.2.4.2	Alex Net	41
2.3.2.4.3	ZFNet	41
2.3.2.4.4	GoogLeNet	42

2.3.2.4.5	VGGNet . . . . .	42
2.3.2.4.6	ResNet . . . . .	43
2.3.2.4.7	FayNet . . . . .	44
<b>2.4</b>	<b>Treinamento . . . . .</b>	<b>45</b>
<b>2.5</b>	<b>Ferramentas de Validação do modelo . . . . .</b>	<b>46</b>
2.5.1	Validação cruzada K-fold . . . . .	46
2.5.2	Métricas de desempenho . . . . .	46
2.5.3	Curva Característica de Operação do Receptor (ROC) e Área Sob a Curva (AUC) . . . . .	48
<b>3</b>	<b>METODOLOGIA . . . . .</b>	<b>51</b>
<b>3.1</b>	<b>Banco de dados . . . . .</b>	<b>51</b>
<b>3.2</b>	<b>Pré-processamento . . . . .</b>	<b>53</b>
<b>3.3</b>	<b>Redes a serem testadas . . . . .</b>	<b>54</b>
3.3.1	FayNet . . . . .	55
3.3.2	ResNet18 . . . . .	56
<b>3.4</b>	<b>Avaliação de desempenho . . . . .</b>	<b>59</b>
<b>4</b>	<b>RESULTADOS E DISCUSSÕES . . . . .</b>	<b>61</b>
<b>4.1</b>	<b>Pré-processamento . . . . .</b>	<b>61</b>
<b>4.2</b>	<b>FayNet . . . . .</b>	<b>62</b>
<b>4.3</b>	<b>Resnet . . . . .</b>	<b>69</b>
<b>5</b>	<b>CONCLUSÃO . . . . .</b>	<b>73</b>
	<b>REFERÊNCIAS . . . . .</b>	<b>75</b>



# 1 Introdução

## 1.1 Contextualização

A esquizofrenia é um dos mais severos transtornos mentais, que afeta cerca de 1% da população mundial. É um transtorno de difícil caracterização e que necessita para isso, de um grande número de análises clínicas que avaliam parâmetros como delírios, alucinações, discursos desorganizados, distúrbios de pensamento dentre diversos outros necessários para realização do diagnóstico (KHODAEI; HOSSEIN-ZADEH; ANANLOO, 2015; SILVA et al., 2016).

Os sintomas referentes à esquizofrenia apresentam uma certa heterogeneidade que permite a visualização desses sintomas em 3 grupos de acordo com suas características. São eles, sintomas positivos, em que a identificação do transtorno é realizada facilmente, uma vez que as características são evidentes (alucinações e delírios), sintomas negativos, onde há uma certa dificuldade de expressar emoções (apatia e alogia) e sintomas cognitivos (deficit de atenção, memória e disfunção executiva). Os pacientes que apresentam características dos grupos de sintomas negativos e cognitivos possuem um maior grau de complexidade em seu diagnóstico, e seus sintomas frequentemente são confundidos com outras doenças, fazendo com que o paciente receba tratamentos ineficazes. (DEMILY; FRANCK, 2008).

A esquizofrenia impacta consideravelmente o bem estar de seu portador e de seus familiares, uma vez que a doença ocasiona mudanças significativas na vida do indivíduo (SOUZA; COUTINHO, 2006). Uma das maiores dificuldades da esquizofrenia reside em sua cronicidade e na recorrência dos sintomas agudos. A apresentação desses dois tipos de sintomas é responsável pelas consequências pessoais do convívio com a esquizofrenia que são refletidas na vida afetiva, social, familiar e financeira, com destruição de sonhos, desconsideração do portador como ser humano, sentimentos de menos valia e sentimento de não ser compreendido pelos familiares e amigos (OLIVEIRA; FACINA; JÚNIOR, 2012; SILVA et al., 2016).

## 1.2 Definição do problema

A esquizofrenia é diagnosticada com base em uma grande coleção de observações clínicas e sintomas psiquiátricos. É uma síndrome difícil de diagnosticar e definir claramente. Em sua maioria, o método de diagnóstico é feito a partir do estudo dos sintomas e requer a experiência de um médico treinado. Além desse método subjetivo de diagnós-

tico, ainda há pouca pesquisa sobre o desenvolvimento de novos métodos objetivos para quantificar os sintomas da esquizofrenia ([CHAKRABORTY et al., 2017](#)).

De forma subjetiva, o médico especialista entrevista o paciente e familiares ou amigos próximos, e o diagnóstico só é levado em conta quando o indivíduo possui características do grupo de risco. Para verificação de quem se encaixa nesse grupo são considerados como parâmetros fatores comportamentais como, por exemplo, a capacidade de socialização ([ASSOCIATION et al., 2014](#)).

No entanto, embora esse método de diagnóstico de esquizofrenia seja aceito ainda hoje com resultados aceitáveis, a demora que ocorre durante o processo pode ser grande, durando muitos anos, e afetar a possível recuperação do paciente devido a demora na identificação da doença ([KIRCHER et al., 2018](#)).

Considerando a fragilidade no método de diagnóstico utilizado atualmente, a busca por métodos mais eficazes se torna relevante. Nesse contexto, uma solução que vem ganhando espaço na área da biomédica é a utilização de neuroimagens para análise de alterações cerebrais em distúrbios mentais ([VERGARA, 2018](#); [NOGUEIRA, 2019](#); [SHARMA; RAMKIRAN, 2015](#)).

### 1.3 Justificativa

A esquizofrenia é uma doença psiquiátrica crônica que perturba e desabilita cerca de 1% da população mundial, que pode ocasionar dependências por parte do paciente devido aos seus sintomas, caso não ocorra seu devido tratamento. Estes sintomas incluem delírios e alucinações, que são precursores de atos de tentativas de suicídio, um dos efeitos correlacionados aos pacientes diagnosticados com a doença ([ASSOCIATION, 2013](#)).

Esta doença ainda é um transtorno mental de difícil compreensão, apesar das diversas análises sobre o assunto ([SILVA et al., 2016](#)). Assim, é possível observar a necessidade de um maior esforço para encontrar métodos mais eficientes para auxiliar no diagnóstico de esquizofrenia, buscando reduzir o tempo atual necessário, podendo melhorar assim a eficácia do tratamento desses pacientes.

A classificação automática utilizando aprendizado de máquinas se torna nesse contexto uma ferramenta com alto potencial. Assim, a busca por arquiteturas com bons resultados em tarefas de classificação similares torna-se viável para encontrar as técnicas que melhor se adaptam à tarefa, como as redes neurais no aprendizado profundo.

## 1.4 Objetivos

### 1.4.1 Objetivos Gerais

Fazer uma análise entre os métodos de classificação automática, comparando diferentes arquiteturas de redes neurais convolucionais, utilizando transferência de conhecimento, ou seja, buscando-se arquiteturas com bons resultados em tarefas de classificação similares, para serem utilizadas na classificação de pacientes em dois grupos, esquizofrênicos e controles saudáveis, a partir das imagens de ressonância magnética (ou do inglês MRI, *magnetic resonance imaging*) do cérebro disponibilizadas pelo *SchizConnect*.

### 1.4.2 Objetivos Específicos

- Realizar o levantamento de técnicas de redes neurais convolucionais para a classificação de MRI.
- Selecionar arquiteturas que apresentaram bons resultados em tarefas similares de classificação que sejam aplicáveis a MRI's do cérebro.
- Realizar a comparação e a análise das arquiteturas propostas através de métricas de desempenho na classificação de esquizofrenia.

## 1.5 Estrutura da Dissertação

A organização do restante deste documento será da seguinte forma. No capítulo 2 trazemos alguns conceitos e revisões necessárias sobre a esquizofrenia e seu diagnóstico, técnicas de neuro imageamento e aquisições de imagens cerebrais, como MRI. O capítulo também apresenta a teoria de ferramentas de processamento de imagens, técnicas de aprendizado de máquina para classificação automática e discussões sobre a esquizofrenia e as redes neurais convolucionais. No capítulo 3 serão apresentados todos os métodos e técnicas utilizadas para alcançar os objetivos do trabalho, assim como a exibição do banco de dados e a descrição necessária para a implementação dos modelos. Os resultados e as discussões obtidos pelos modelos propostos na metodologia estarão expostos no capítulo 4, no qual trará comparações entre as CNN's discutidas. O documento será finalizado no capítulo 5 de modo a tratar as conclusões da dissertação, expondo as suas contribuições geradas e de possíveis sugestões de estudo para trabalhos futuros.



## 2 Referencial teórico

### 2.1 Esquizofrenia

Segundo *Schizophrenia Working Group of the Psychiatric Genomics Consortium* ([CONSORTIUM, 2015](#)), a esquizofrenia é caracterizada por alucinações, perda de contato com a realidade, delírios, variação emocional, disfunção ocupacional e social, além do comprometimento do raciocínio, solução de problemas, discurso e comportamento que se apresentam desorganizados.

Os sintomas de forma geral começam a ser apresentados na adolescência ou no início da fase adulta e mostram-se em circunstâncias que duram até 6 meses antes de seu diagnóstico. Seu princípio é uma incógnita, porém há vestígios de algum componente genético e ambiental. A detecção e o diagnóstico imaturo indicam um melhor resultado por melhorar e facilitar o tratamento, que é feito com medicamentos, terapia cognitiva e acompanhamento psiquiátrico ([CONSORTIUM, 2015](#)).

#### 2.1.1 Diagnóstico

Para o diagnóstico da esquizofrenia, não existe um teste definitivo. O diagnóstico é fundamentado no histórico clínico, sintomas e sinais do paciente, além de informações de seus familiares, amigos e outros indivíduos de convívio próximo. De acordo com *Diagnostic and Statistical Manual of Mental Disorders, fifth edition – (DSM – 5)* ([ASSOCIATION et al., 2014](#)), o diagnóstico de esquizofrenia demanda de dois aspectos:

- Possuir dois ou mais sintomas característicos com atuações significativas por um período 6 meses;
- Sinais de indícios iniciais de enfermidades com prejuízos sociais, ocupacionais ou de cuidados pessoais devem ficar evidentes por período de 6 meses, incluindo 1 mês de sintomas ativos.

Os sintomas para o diagnóstico são ([CLEMENTZ et al., 2016](#)):

- Delírios;
- Alucinações;
- Fala desorganizada;
- Comportamento desorganizado;

Estes são ditos como sintomas positivos, pois são indícios apresentados de forma nítida além de atuar de maneira mais presente.

Há ainda os sintomas negativos, que trazem a perda ou redução de motivação e emoção, onde o paciente desabitua de seus propósitos e objetivos, além da perda de interesses pessoas em algumas atividades que antes era satisfatório fazer. Estes sintomas podem ser descritos como (CLEMENTZ et al., 2016):

- Pobreza de fala;
- Anedonia;
- Falta de sociabilidade;
- Afeto embotado;

Esses sintomas podem estar presentes em diversos transtornos psicóticos, no entanto é possível diferenciar um transtorno de outro através da duração e quantidade de cada sintoma (ASSOCIATION et al., 2014). Uma outra alternativa para o diagnóstico desta doença é a análise morfológica do cérebro, onde é apresentada anormalidades em sua estrutura referente a características específicas do transtorno (KHODAEI; HOSSEIN-ZADEH; ANANLOO, 2015).

### 2.1.2 Neuro Imageamento

Alterações cerebrais desenvolvidas por pacientes portadores de esquizofrenia estão sendo cada vez melhores visualizadas, de acordo com o aprimoramento dos métodos de aquisição de imagens. A tomografia computadorizada (ou do inglês, CT Computed tomography) e imageamento por ressonância magnética são as principais ferramentas utilizadas por possuírem melhor qualidade para análise (TORRES, 2016). O uso destas imagens é apresentada regularmente para o auxílio de diagnósticos, pois é possível comparar as anormalidades anatômicas do cérebro entre pacientes selecionados, como um volume em uma dada região de interesse ou o tamanho de porções do cérebro (KHODAEI; HOSSEIN-ZADEH; ANANLOO, 2015).

#### 2.1.2.1 Imagens de ressonância Magnética - MRI

A ressonância magnética é uma tecnologia de imagem não invasiva que produz imagens anatômicas detalhadas tridimensionais. É frequentemente usada para detecção, diagnóstico e monitoramento no tratamento de doenças. Baseia-se em uma tecnologia sofisticada que estimula e detecta a mudança na direção do eixo de rotação dos prótons encontrados na água que compõem os tecidos vivos (IMAGING; NIBIB, 2018).

A MRI apresenta algumas vantagens nos detalhes anatômicos em comparação a outras tecnologias para a produção deste tipo de imagem, por não utilizar qualquer tipo de radiação ionizante, ou qualquer tipo de intervenção danosa ao paciente, como o caso da tomografia computadorizada. Devido a isso, o uso de MRI se tornou popular entre os centros de pesquisas e de diagnósticos, sendo que o primeiro estudo envolvendo pacientes portadores de esquizofrenia utilizando essa técnica de imageamento foi realizado por Smith em 1984 (SMITH et al., 1984; FERRARI, 2008);

A MRI é a técnica de imageamento médico que mais evoluiu na última década em comparação com outras técnicas existentes. A MRI se destaca pelo seu grande contraste, sendo possível diferenciar diversos tipos de tecido, órgãos e identificar mudanças anatômicas com alta precisão (CAMPINAS, 2015).

### 2.1.3 Alterações cerebrais observadas em pacientes portadores de esquizofrenia

Vários estudos evidenciam que há diferenças perceptíveis nas regiões cerebrais de indivíduos portadores de esquizofrenia comparados com indivíduos saudáveis (KHODAEI; HOSSEIN-ZADEH; ANANLOO, 2015; ANDREASEN et al., 2011). O trabalho de Van Erp (ERP et al., 2016) trás algumas informações sobre as alterações que ocorrem no cérebro de pacientes portadores de esquizofrenia comparados a pacientes de um grupo de controle saudáveis, utilizando como ferramenta para isso imagens de ressonância magnética do cérebro. Dentre as alterações observadas, as mais perceptíveis foram a redução do tamanho do hipocampo, um aumento no volume do pálido e aumento do ventrículo lateral.

Como uma das alterações mais perceptíveis ao se analisar imagens de ressonância magnética de pacientes portadores de esquizofrenia é o aumento do ventrículo, diversos trabalhos vêm estudando maneiras de melhorar o diagnóstico da doença. Para isso a utilização do aprendizado de máquinas está se demonstrando cada vez mais eficiente, otimizando tarefas como, segmentação e classificação de pacientes portadores de esquizofrenia a partir de MRI cerebral (TORRES, 2016; VERGARA, 2018; NOGUEIRA, 2019).

## 2.2 Processamento digital aplicado a imagens

A utilização do processamento de imagens juntamente com o aprendizado de máquinas têm se demonstrado uma ótima maneira de extrair características de imagens e realizar a classificação entre grupos, utilizando ferramentas como redes neurais convolucionais (ou do inglês, *CNN Convolutional Neural Network*) e máquinas de vetor de suporte (ou do inglês, *SVM Support Vector Machine*). Diversos trabalhos têm demonstrado a efi-

ciência na utilização dessas ferramentas em aplicações semelhantes (ALFONSE; SALEM, 2016; MATHUR; JAIN; SINGH, 2017; SAYED; ZAGHLOUL; NASSEF, 2016).

### 2.2.1 Processamento digital de imagens

Para entender um pouco do conceito de processamento digital de imagens, primeiro é necessário uma definição do que caracteriza uma imagem. Ela pode ser definida como uma função bidimensional,  $f(x, y)$  em que  $x$  e  $y$  são coordenadas espaciais (plano), e a amplitude de  $f$  em qualquer par de coordenadas  $(x, y)$  é chamada de intensidade ou nível de cinza da imagem nesse ponto. Quando  $x$ ,  $y$  e os valores de intensidade de  $f$  são quantidades finitas e discretas, chamamos de imagem digital. O campo do processamento digital de imagens se refere ao processamento de imagens digitais por um computador digital. Observe que uma imagem digital é composta de um número finito de elementos, cada um com localização e valor específicos, sendo "*pixel*" (do inglês *picture element*) o menor elemento constituinte da imagem (GONZALEZ; WOODS; EDDINS, 2004).

A partir desses aspectos das imagens digitais, são possíveis realizar diversos processamentos que ajudam em tarefas relevantes, como a segmentação de tecidos, extração de características, classificação de imagens dentre diversos outros. Na classificação automática utilizando redes neurais convolucionais, por exemplo, pode haver a necessidade de redimensionar as imagens a fim de adequá-las ao tamanho especificado na construção do classificador. Esse redimensionamento pode ser feito através de algoritmos que fazem a exclusão de algumas linhas e colunas de *pixels* da imagem, quando o intuito é reduzi-la e a partir de métodos de interpolação, onde há o aumento de linhas e colunas com base em cálculos estatísticos dos *pixels* vizinhos.

Existem ainda alguns *softwares* que possuem opções de pré-processamento de imagens que já vem com algoritmos de manipulação embutidos e que trazem uma maior facilidade na realização do pré-processamento como, por exemplo o software CLÍNICA ([www.clinica.run](http://www.clinica.run)), uma plataforma de estudos de pesquisa em neurociência dedicada a análise de neuroimagens.

### 2.2.2 Convolução de imagens digitais

A convolução consiste em uma operação linear que, a partir de duas funções  $x$  e  $h$  resulta numa terceira  $y$ , onde  $x$  é a imagem original,  $h$  é a resposta ao impulso do filtro a ser utilizado e  $y$  é o resultado desse processo.

No processamento de imagens, a convolução realiza esse processo sobrepondo a imagem com uma máscara (chamada de *kernel*), movendo-a um *pixel* por vez pela imagem, fazendo o cálculo de somatório do produto dos *pixels* sobrepostos. A Figura 1, ilustra esse processo considerando uma imagem 2D 6x6 e um filtro 2x2 a ser convoluído pela



imagem. Outra forma de visualizar esse processo é observando a equação 2.1, que descreve analiticamente esse funcionamento.

$$y(n1, n2) = \sum_{k1=0}^N \sum_{k2=0}^M x(k1, k2).h(n1 - k1, n2 - k2), \quad (2.1)$$

onde,  $N$  e  $M$  são a altura e largura das imagens  $h$  é a resposta ao impulso do filtro,  $x$  é a imagem original e  $y$  é a imagem resultante da convolução do filtro por essa imagem.

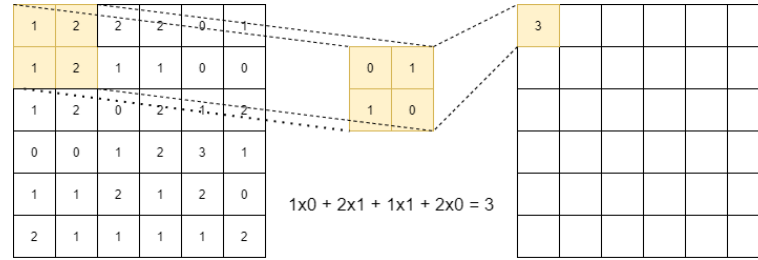


Figura 1 – Exemplo de uma operação de convolução de uma imagem 6x6 por um filtro 2x2. Fonte: Autor

## 2.3 Aprendizado de Máquinas

Descobrir padrões em sinais vêm sendo um tópico de interesse por grande parte dos pesquisadores por muito tempo (BISHOP, 2006). E, tratando-se de imagens digitais, o aprendizado de máquina vem se mostrando uma ferramenta altamente eficaz para realizar o reconhecimento desses padrões.

Esse aprendizado é feito a partir da execução de um programa de computador para otimizar os parâmetros do modelo usando os dados de treinamento ou experiência passada. O modelo pode ser preditivo (fazendo previsões no futuro), descritivo (obtendo conhecimento dos dados) ou ambos, sendo estes representados por funções matemáticas provenientes da teoria estatística, uma vez que sua tarefa principal é fazer inferências a partir de amostras (ALPAYDIN, 2016).

### 2.3.1 Classificação automática aplicada a imagens

A classificação consiste em encontrar características da imagem, e de acordo com sua similaridade, decidir se os padrões encontrados são predominantes de uma determinada classe. Um dos maiores desafios é construir um bom modelo a partir do conjunto de características de entrada do algoritmo, e usá-las para reconhecimento de padrões (GONZALEZ; WOODS; EDDINS, 2004). Dentre os modelos de classificação, os mais usados são os modelos de aprendizado supervisionado e modelo de aprendizado não-supervisionado.

Na aprendizagem supervisionada, um conjunto de dados de entrada são previamente rotulados por um “supervisor”. Nesse aprendizado é necessário ainda fornecer uma

função, que busca fazer a distinção entre os dados ainda não rotulados com base nas amostras conhecidas. Já na aprendizagem não supervisionada, os algoritmos experimentam um conjunto de dados contendo muitos recursos e, em seguida, aprendem propriedades úteis referente a estrutura desse conjunto de dados. Alguns dos algoritmos de classificação mais comuns da aprendizagem tradicional são as SVM, KNN e *Random Forest*.

A partir dos métodos de aprendizagem tradicional apresentados, nota-se que para o sucesso destes classificadores são necessárias técnicas preliminares de processamento de sinais. Um ponto negativo está relacionado à extração de características do sinal de entrada. Este tipo de extração é uma parte fundamental do problema, em que um especialista da área necessita de total domínio do problema para encontrar as melhores características a serem utilizadas (VERGARA, 2018). O aprendizado profundo é uma ótima alternativa para resolver o problema da extração de características dos algoritmos de aprendizado de máquinas tradicionais, pois é uma técnica desenvolvida para extrair e manipular características de dados brutos.

No contexto de aprendizado profundo, geralmente é desejado que se aprenda toda a distribuição de probabilidade que gerou um conjunto de dados, seja explicitamente, como na estimativa de densidade, ou implicitamente, para tarefas como síntese ou redução de ruído (GOODFELLOW et al., 2016).

### 2.3.2 Redes Neurais artificiais - RNA

As redes neurais artificiais ou RNA são sistemas de processamento de dados com o objetivo de distingui-los para saídas desejadas. Tais sistemas recebem esse nome por serem inspiradas no funcionamento do córtex cerebral dos mamíferos. Porém, é importante notar que não se assemelham a complexidade, funcionamento e escala do cérebro humano. A arquitetura de uma RNA é composta por um grupo de camadas de neurônios ou nós organizados em uma estrutura que possui uma conexão paralela sólida, de modo a avaliar os dados de entrada e saída. Assim, as saídas serão resultantes da interconexão dos dados de entrada, em que a informação é recebida e distribuída pelos neurônios através de recursos de ativação, sendo tratadas por um peso durante a comunicação entre os nós (KHAN et al., 2018; EFFENDY et al., 2020).

De uma maneira simples, os modelos de redes neurais artificiais podem ser compreendidos como um tratamento de um grupo de dados que estão interconectados, fornecidas como entrada, para processar suas informações e gerar saídas esperadas. Existem duas categorias em que estes modelos podem ser distribuídos, o *feed-forward* e o *feed-back*, sendo agrupados de acordo como as informações propagadas na rede. O fluxo de informações em uma rede *feed-forward* ocorre apenas em uma direção. Um exemplo de arquitetura desta categoria é a Rede Neural Convolutiva (CNN). Já as redes de *feed-back* possuem conexões que formam ciclos direcionados e a capacidade de memorização de informações

em sua memória interna. As arquiteturas de Rede Neural Recorrente (RNN) e Memória de Longo-Curto Prazo (LSTM), são inclusas nessa categoria (KHAN et al., 2018).

### 2.3.2.1 Redes de camada única

Na literatura, existem diversos tipos de redes neurais para várias aplicações. As redes neurais de camada única são a arquitetura mais simples dentre as redes e por isso é comum sua utilização (HAYKIN, 2007). Essa rede é um sistema *feed-forward* apresentada por Perceptrons de Multicamadas (ou do inglês, MLPs *Multilayered Perceptrons*), por possuir uma camada de entrada, uma camada de saída e uma única camada oculta ou intermediária, no qual esta presente o neurônio que realiza o processamento da rede. Esse modelo, exibido na Figura 2, foi desenvolvido por McCulloch e Pitts em 1943 (MCCULLOCH; PITTS, 1943) e está limitado a classificação de padrões de duas classes.

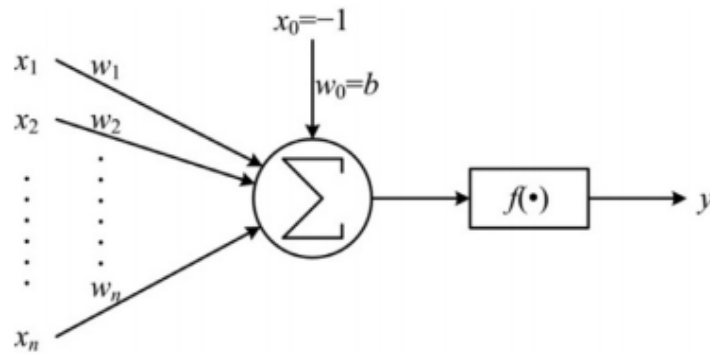


Figura 2 – Modelo de um *perceptron*. Fonte: (ZHANG et al., 2020)

Na Figura 2, as entradas do neurônio são dado por  $\{x_1, x_2, \dots, x_n\}$ ,  $y$  é a saída do sistema,  $\{w_1, w_2, \dots, w_n\}$  são os pesos,  $b$  o limite e  $f$  a função de ativação. O modelo pode ser representado por

$$y = f\left(\sum_{i=1}^n w_i x_i - b\right). \quad (2.2)$$

### 2.3.2.2 Deep learning e Redes de múltiplas camadas

Aprendizado profundo ou do inglês *Deep learning*, é um método para se executar o aprendizado de máquina, em que os computadores formam grandes redes neurais artificiais inspiradas nas redes neurais biológicas humanas (HINTON; SHALLICE, 1991).

Há indícios de que os algoritmos de aprendizagem de máquina e classificação automática de sinais tinham a restrição de interpretar dados de entrada completos, sem uma análise prévia das particularidades significativas para o seu processamento. Através dos métodos de aprendizagem profunda, propõem-se distinguir características invariantes do sinal e classificar suas representações abstratas de forma automática, com o intuito de

transformar estas abstrações mais resistentes a possíveis ruídos no sinal de entrada do sistema, tornando o modelo de aprendizado mais robusto e sofisticado (LECUN; BENGIO; HINTON, 2015; LESHNO et al., 1993).

Um exemplo do modelo de aprendizado profundo é a rede profunda *feed-forward* ou Perceptron Multicamadas (MLP do inglês *Multilayer Perceptrons*). Um perceptron multicamadas consiste em uma função matemática que mapeia alguns conjuntos de valores de entrada para valores de saída. Sendo a função formada pela composição de muitas funções mais simples (GOODFELLOW et al., 2016).

Existem várias técnicas para a implementação do aprendizado profundo para cada tipo de aplicação, devendo-se levar em consideração o tipo de dados disponíveis ou se o modelo será supervisionado ou não supervisionado, por exemplo. Nos dias atuais, pode-se observar estas implementações em diversas aplicações, como na área de robótica, reconhecimento de objetos, reconhecimento e análise de padrões, extração de características e classificação. Estas aplicações são possíveis devido ao fato de que a aprendizagem profunda, diferente de arquiteturas rasas, possui múltiplos níveis de aprendizado. Então, a partir de um sinal de entrada, cria-se de uma forma automática e hierárquica diversos modelos multiníveis que em cada estágio e camada da arquitetura, se torna mais abstrato, chegando finalmente na camada final de saída, como o resultado de uma classificação (LECUN; BENGIO; HINTON, 2015; Zhang et al., 2019; ZHANG et al., 2020).

#### 2.3.2.3 Redes Neurais convolucionais - CNN

As CNN's são um tipo especializado de rede neural para processamento de dados e têm sido extremamente bem-sucedidas em aplicações práticas (GOODFELLOW et al., 2016). Essas redes são variantes do conhecido perceptron multicamadas e estão inspiradas biologicamente segundo a arquitetura multi-estado de Hubel-Wiesel (HUBEL; WIESEL, 1962).

Desenvolvida em 1989 por LeCun (LECUN et al., 1989) as redes neurais convolucionais só ganharam uma maior atenção posteriormente com o avanço da tecnologia, no melhoramento de GPU's (ou do inglês, Graphics Processing Unit), fazendo com que se tornasse viável o custo computacional de se utilizar uma rede convolucional. Em 2011 Cirezan e colaboradores (CIRESAN et al., 2011) trouxeram resultados impressionantes ao realizar um aprimoramento das CNN's e realizar a implementação em uma GPU.

A entrada de uma rede neural convolucional, no caso da classificação de imagens, pode ser uma matriz bidimensional ou tridimensional, dependendo do formato da imagem desejada. Como as imagens possuem uma altura, uma largura e uma profundidade elas são também chamadas de "volume de entrada". Internamente a rede é composta por estágios onde cada estágio é composto de três camadas: uma camada de banco de filtros, uma camada de não linearidade e um camada de *pooling* de recursos. Uma rede convolucional

típica é composta por um, dois ou três desses estágios de três camadas, seguidos por um módulo de classificação (LECUN; KAVUKCUOGLU; FARABET, 2010).

A Figura 3 ilustra uma formação básica de uma CNN, considerando como entrada uma imagem bidimensional, onde um filtro de dimensões ( $m \times m$ ) é convoluído pela entrada por  $n$  vezes formando um total de  $n$  kernels para cada uma das camadas.

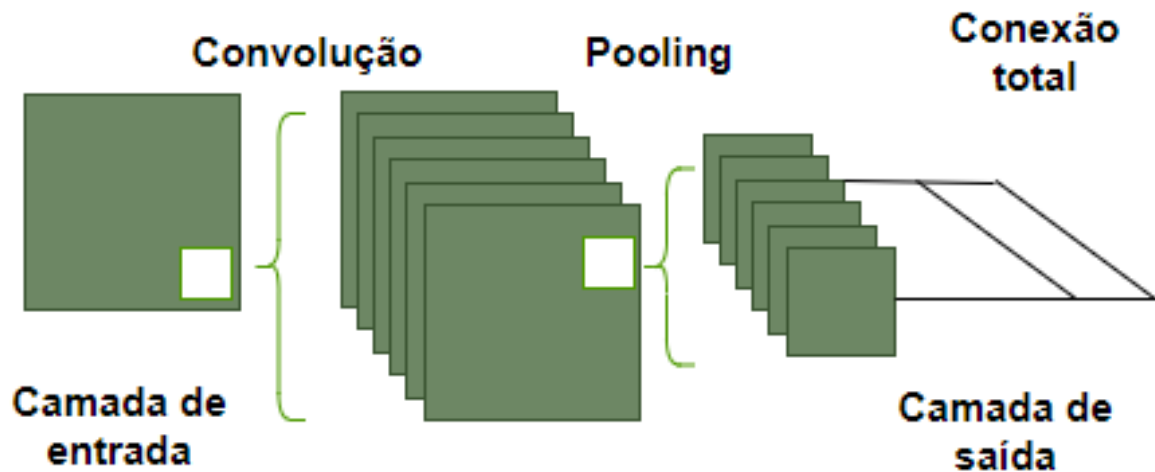


Figura 3 – Arquitetura base de uma CNN. Fonte: Adaptado de (VERGARA, 2018).

#### 2.3.2.3.1 Camada convolucional

Como o próprio nome já diz, as “Redes Neurais Convolucionais” fazem a maior parte do trabalho computacional através da convolução. Seus parâmetros são um conjunto de filtros treináveis. Cada filtro é sensível a uma determinada característica que pode ser, por exemplo, uma borda vertical, diagonal ou horizontal. A saída desses filtros é conhecida como mapa de características. Uma camada convolutiva tem vários mapas de características, produzidos por planos de neurônios, cada plano implementa um filtro diferente. Geralmente, os filtros são pequenos espacialmente (ao longo da largura e altura), mas estende-se através de toda a profundidade do volume de entrada (LECUN; KAVUKCUOGLU; FARABET, 2010).

Quando as entradas possuem altas dimensões, não é viável conectar todos os neurônios no volume anterior. Em vez disso cada neurônio é conectado em uma região local do volume de entrada. A extensão dessa conectividade é um hiper-parâmetro denominado campo receptivo do neurônio. Já a extensão da conectividade ao longo do eixo de profundidade é sempre igual à profundidade do volume de entrada. Dessa maneira cada unidade é sensível a variações fora do seu campo receptivo, garantindo que os filtros aprendidos produzam respostas inerentes aos padrões de entrada de uma dada região espacial (KARPATHY, 2018). Em relação a saída e sua organização, alguns hiper-parâmetros controlam

o tamanho e volume dos neurônios de saída como, por exemplo, o passo do filtro e o preenchimento por zero.

Ao realizar a convolução do filtro com o dado de entrada deve-se definir o passo que o filtro terá para percorrer toda a matriz de entrada, uma vez que se o passo for igual a 1 o filtro é movido um *pixel* de cada vez, quando o passo é 2 os filtros saltam 2 *pixels* por vez e assim sucessivamente. Logo o aumento do valor do passo resulta em um menor volume espacial na saída.

Quando o tamanho do filtro associado com o passo não consegue percorrer toda imagem sem extrapolar seu tamanho original a técnica de preenchimento por zero é aplicada. Essa técnica consiste em adicionar zeros nas bordas das imagens para que o filtro consiga fazer o processo de convolução por toda imagem.

Para calcular a quantidade de neurônios na saída da rede (N) em função do tamanho volume de entrada (W), do tamanho do campo receptivo (F), do passo (S) e da quantidade de preenchimento por zeros (P), pode-se fazer:

$$N = \frac{W - F + 2P}{S} + 1 \quad (2.3)$$

#### 2.3.2.3.2 Camada de Não-linearidades

Após a realização da convolução uma camada não linear é aplicada, uma vez que se deseja deixar o volume da saída da etapa de convolução inalterado. Tradicionalmente em CNN's, isso simplesmente consiste em uma função sigmóide ou a função *tanh()* pontual (LECUN; KAVUKCUOGLU; FARABET, 2010). No entanto, implementações recentes têm buscado novas funções para otimização da convergência da função, sendo uma amplamente utilizada a Unidade Linear Retificante (ou do inglês, ReLU *Rectified Linear Units*).

O trabalho de Nair e Hilton (NAIR; HINTON, 2010) mostra que a velocidade de treinamento utilizando RELU's é maior que outras não linearidades como a *tanh*. É possível visualizar essa diferença a partir da imagem de comparação entre a utilização da RELU e a função *tanh* na Figura 4, onde a RELU apresenta uma velocidade de treinamento 6 vezes maior que a *tanh*.

#### 2.3.2.3.3 Normalização

Ao se realizar o treinamento de redes neurais profundas, o fato da distribuição das entradas de cada camada mudar durante o treinamento, conforme os parâmetros das camadas anteriores mudam, retarda o treinamento. Isso exige taxas de aprendizado mais baixas e inicialização cuidadosa dos parâmetros, tornando notoriamente difícil treinar

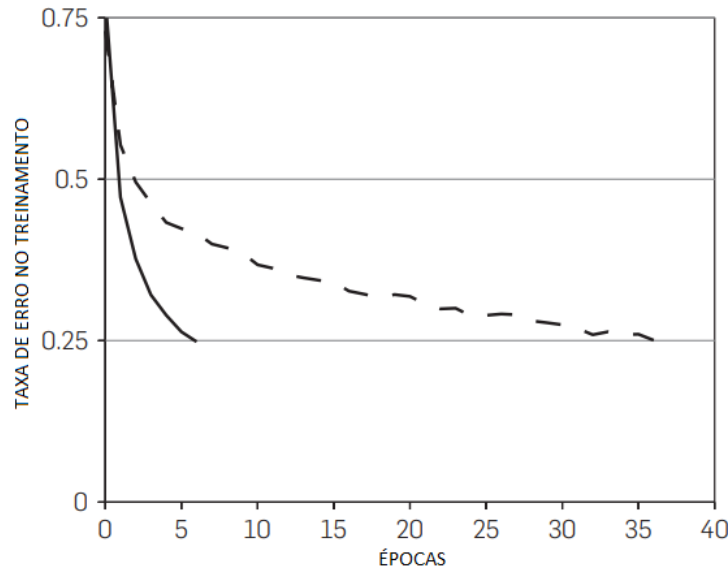


Figura 4 – Comparação entre as funções de perda utilizando ReLU (linha sólida) e tanh (linha tracejada) realizada por Krizhevsky em um problema de classificação. Demonstrando a maior velocidade no treinamento utilizando a RELU. Fonte: (KRIZHEVSKY; SUTSKEVER; HINTON, 2017a)

modelos com não linearidades. Referimo-nos a esse fenômeno como mudança de covariável interna. Esse problema pode ser resolvido normalizando as entradas das camadas a cada lote de treinamento. Com a normalização em lotes, a taxa de aprendizado pode ser então mais alta, e podemos ser um pouco menos cuidadosos com a inicialização dos parâmetros (IOFFE; SZEGEDY, 2015).

#### 2.3.2.3.4 Camada de Pooling

Para reduzir a quantidade de amostras, uma etapa de subamostragem é necessária. Essa etapa é chamada de *pooling*, onde busca-se resumir as saídas de grupos vizinhos de neurônios no mesmo mapa de características (GOODFELLOW et al., 2016). É comumente inserida entre as camadas sucessivas de convolução, para assim reduzir progressivamente o tamanho espacial e reduzir a quantidade de parâmetros, controlando também o sobreajuste (KARPATHY, 2018). Uma das operações de *pooling* mais comuns para realizar o redimensionamento é o *pooling* máximo geralmente abreviado como *max-pooling*.

O *max-pooling* (ZHOU et al., 1988) consiste em tomar o valor máximo em uma determinada janela. No entanto, existe também o *Average Pooling*, onde é feito uma média dos valores da janela para se determinar o novo valor.

O processo de redimensionamento está ilustrado na Figura 5 com o *max-pooling*. Neste exemplo é considerada uma entrada com volume  $W \times H \times D$ , uma camada de *max-pooling* com filtro  $2 \times 2$  e um passo de 2.

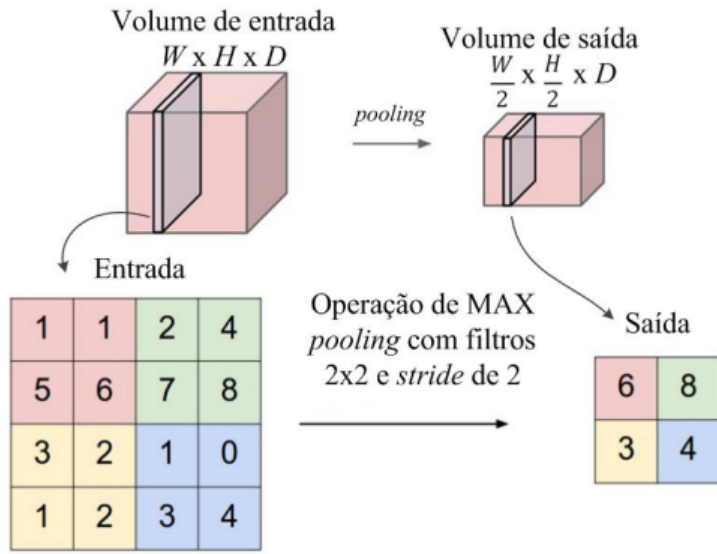


Figura 5 – Filtro 2x2 deslocando-se na imagem com passo de 2, de forma que entre os 4 elementos da janela é escolhido o de maior valor. Fonte: Adaptado de (KARPATHY, 2018).

#### 2.3.2.3.5 Camada Totalmente Conectada

A camada totalmente conectada é inserida ao final do modelo de aprendizado profundo, com o intuito de avaliar a qualidade da rede, a partir de cálculos de probabilidade de acertos para a tarefa de classificação desejada. Para realizar a comparação entre o resultado final na saída da rede e o rótulo designado para o dado na entrada é utilizando uma função de perdas.

##### ***Softmax***

Na literatura, o modelo de decisão mais utilizado para problemas multi-classe é a função de ativação *softmax*, que pode ser representada pela regressão logística clássica de classificação binária (Reverdy; Leonard, 2016; HAN et al., 2020), dada por

$$Pr(i) = \frac{e^{x_i}}{\sum_{k=1}^K e^{x_k}}, \quad (2.4)$$

no qual define o rótulo  $k \in \{0, 1\}$  através dos recursos de entrada  $x$ . Em outras palavras, o *softmax* propõe mapear os recursos de entrada na probabilidade da classe  $k$ , em que o somatório de todas as probabilidades é igual a 1. (HAN et al., 2020).

##### ***Sigmoid***

A função *sigmoid* possui sua maior vantagem pelo fato de que sua derivada é máxima quando  $x$  está próximo de 0, isso faz com que esta função tenda a jogar o resultado para as extremidades do intervalo  $[0, 1]$  ao longo do treinamento, o que é desejável em problemas de classificação, como por exemplo, problemas de classificação binários (GOMES; LUDERMIR, 2008).



A função sigmoide ou logística tem seu comportamento descrito pela equação 2.5.

$$\sigma(x) = \frac{1}{1 + e^x}. \quad (2.5)$$

#### 2.3.2.3.6 Camada de *Dropout*

Redes neurais profundas contêm várias camadas ocultas e não lineares fazendo com que muitos modelos expressivos possam aprender relações muito complicadas entre suas entradas e saídas. Muitos dessas conexões complexas são resultado da amostragem de ruído. Assim eles existirão no conjunto de treinamento, mas não na realidade dos dados de teste, levando ao chamado sobre-ajuste (do inglês, *overfitting*). Uma característica de um modelo de rede que apresenta *overfitting* é a capacidade de obter resultados com altíssima precisão quando testado com seu conjunto de dados, porém com dados reais o resultado apresentado não possui precisão correspondente aos testes (SRIVASTAVA et al., 2014).

O *dropout* (HINTON et al., 2012) é um método de regularização para evitar o *overfitting* que tem sido empregado recentemente, e é muito utilizado na aprendizagem profunda (ROMAGUERA et al., 2017; VERGARA, 2018) a fim de evitar a produção de co-adaptações com os dados de treinamento. Esta técnica consiste em zerar a saída de cada neurônio da camada de aprendizado com uma certa probabilidade, portanto não contribuindo para a sequência de classificação da arquitetura. A Figura 6 demonstra o funcionamento desse método.

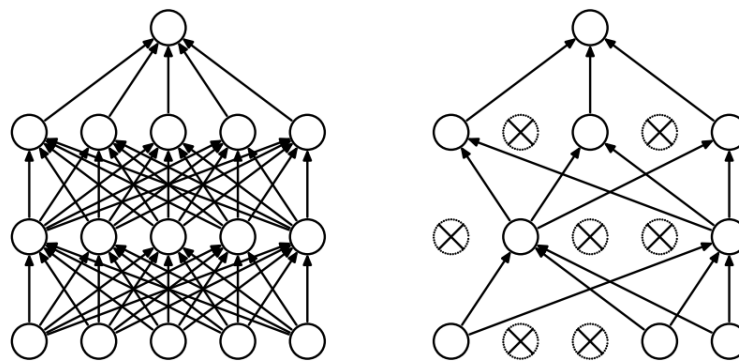


Figura 6 – À esquerda: Uma rede neural padrão com 2 camadas ocultas. À direita: um exemplo de rede reduzida produzida pelo *dropout*. Unidades cruzadas foram descartadas. Fonte: (SRIVASTAVA et al., 2014)

#### 2.3.2.4 Tipos de CNN's

As CNNs fazem parte de uma família *feed-forward* de redes profundas, em que as camadas intermediárias recebem como entrada os recursos gerados pela camada anterior, e passam suas respectivas saídas para a camada subsequente. Existem várias arquiteturas para a implementação de uma CNN, em que o sucesso destas redes está no aprendizado de

camadas hierárquicas de representação de conceitos, correspondentes a diferentes níveis de abstração (SAIKIA et al., 2019). Assim estas arquiteturas se diferenciam devido a sua complexidade em camadas mais profundas, como a quantidade de filtros por camadas e camadas convolucionais empilhadas, por exemplo. A arquitetura de uma CNN está relacionada de acordo com o problema que será abordado, com isso diversos estudos desenvolveram variadas arquiteturas CNN para a classificação. Entre os mais populares estão: a LeNet, AlexNet, ZFNet, GoogLeNet, VGGNet e ResNet (BORA et al., 2020).

#### 2.3.2.4.1 LeNet

A LeNet, desenvolvida por LeCun em 1998 foi a primeira rede convolucional desenvolvida e testada com êxito (LECUN; K.; C., 2010; Lecun et al., 1998). A rede proposta tinha 7 níveis e seu propósito era a classificação de dígitos escritos à mão. A entrada da rede eram imagens em tons de cinza com tamanho 32x32 pixels e sua arquitetura consistia em dois conjuntos de convoluções e camadas de *pooling* médias, uma camada convolucional achatada, seguidas por duas camadas totalmente conectadas e uma camada *softmax* para a realização da classificação (Lecun et al., 1998). A estrutura da arquitetura desta rede é ilustrada na Figura 7, que foi usado para o reconhecimento de dígitos. Em cada plano é representado como um mapa de características, ou seja, um conjunto de unidades cujos pesos são específicos a serem idênticos (KRIZHEVSKY; SUTSKEVER; HINTON, 2012).

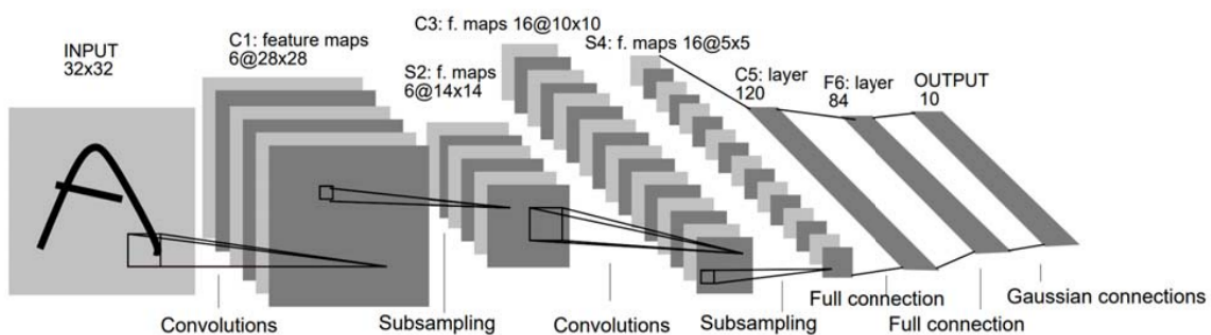


Figura 7 – Arquitetura da rede LeNet. Fonte: (Lecun et al., 1998).

Por ser pioneira e a mais simples das demais, esta técnica é limitada pela disponibilidade de recursos de computação, pois para imagens de resolução mais alta, é necessária camadas maiores e mais profundas (possuindo mais convoluções) (GUA et al., 2017). Porém, possibilitou caminhos para modelos mais complexos e melhores, além de servir como base até hoje.

#### 2.3.2.4.2 Alex Net

Apresentada de modo significativo em 2002, a AlexNet resume-se em 25 camadas com entrada de tamanho aceito de  $227 \times 227 \times 3$ , desenvolvida para o reconhecimento de objetos da ImageNet, um grande banco de dados visual projetado para uso em reconhecimento visual de objetos (KRIZHEVSKY; SUTSKEVER; HINTON, 2017b). A AlexNet é constituída por 5 camadas de convolução seguidas por uma camada ReLU e uma camada de *pooling* máximo. As duas primeiras camadas de convolução dão sequência à camada de normalização do canal cruzado, na qual realiza a normalização por canal, alterando-se cada componente por um valor normalizado proveniente de células vizinhas. A arquitetura ainda possui três camadas totalmente conectadas, em que nas duas primeiras são seguidas por camadas de exclusão, prosseguido por uma camada ReLU. A saída da última camada totalmente conectada é a entrada do *softmax* que produz a distribuição de probabilidade entre 1000 classes (GUYON et al., 1994; KRIZHEVSKY; SUTSKEVER; HINTON, 2017b). Uma representação desta arquitetura está apresentada na Figura 8, onde é possível observar as interações entre as duas GPUs. A primeira GPU executa as partes da camada no topo da figura enquanto que a outra executa as partes da camada no fundo, comunicando-se apenas em algumas camadas.

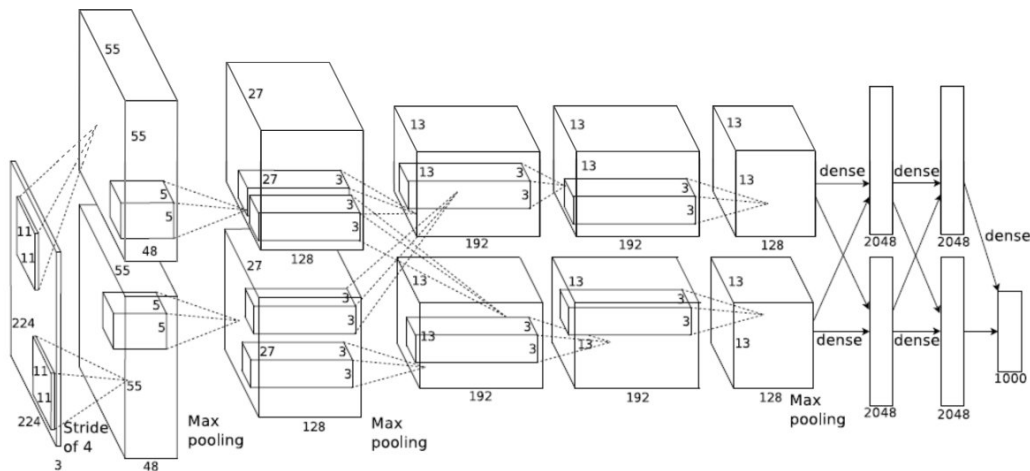


Figura 8 – Arquitetura da rede Alex Net. Fonte: (KRIZHEVSKY; SUTSKEVER; HINTON, 2012).

#### 2.3.2.4.3 ZFNet

De maneira semelhante à arquitetura da AlexNet, a ZFNet, elaborada em 2013 com o mesmo objetivo de reconhecimento de objetos na ImageNet, se difere na quantidade, no tamanho e no passo dos filtros utilizados. O *stride* é um parâmetro de distância no qual é deslocado o filtro (Antioquia et al., 2018). Quando o *stride* é 1, movemos os filtros um pixel de cada vez. Quando o *stride* é 2 (3 ou mais), os filtros saltam 2 pixels por vez, conforme os movemos. Na AlexNet, sua primeira camada de convolução usa um filtro de

tamanho  $11 \times 11$  com *stride* de 4 já na ZFNet usa-se um filtro  $7 \times 7$  e *stride* de 2, ou seja, os filtros saltam 2 pixels por vez, conforme os movemos. Isso produzirá volumes menores de saída espacialmente (UNIVERSITY, 2019). Outra diferença é que na ZFNet é usado filtros de 512, 1024 e 512 na terceira, quarta e quinta camada de convolução enquanto na AlexNet opera-se com filtros  $384 \times 384$  e 256, respectivamente (Antioquia et al., 2018).

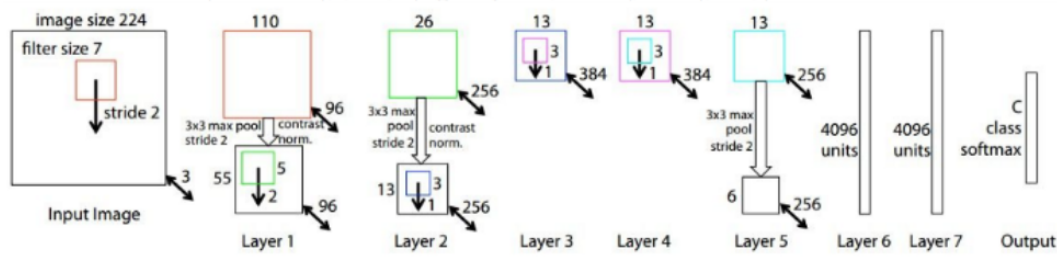


Figura 9 – Arquitetura da rede ZFNet. Fonte: (DAS, 2017).

#### 2.3.2.4.4 GoogLeNet

Diferente das arquiteturas apresentadas, a rede das versões GoogLeNet possuem uma peculiaridade. A rede GoogLeNet foi inspirada na arquitetura LeNet, porém utiliza uma técnica dada por módulo de iniciação, que apresentam diferentes tamanhos de convoluções para a mesma entrada e delimitando todas as saídas, com o intuito de reduzir consideravelmente o número de parâmetros. Assim, a cada nível introduzido, denominado módulo de iniciação, é realizado um aumento na profundidade da rede (SURIANI et al., 2019).

#### 2.3.2.4.5 VGGNet

A rede VGGNet foi desenvolvido por Simonyan e Zisserman (SIMONYAN; ZISSERMAN, 2014), e é apresentada em 16 camadas convolucionais. Uma curiosidade desta arquitetura, conforme ilustrado na Figura 10, é a sua uniformidade. Similar á AlexNet, são usadas apenas convoluções de tamanho  $3 \times 3$ , porém com muitos filtros. Nos dias atuais, tem sido uma das mais escolhidas para a extração de recursos em imagens. No entanto, a VGGNet consiste em 138 milhões de parâmetros, o que pode ser um pouco desafiador de manipular (DHOMNE; KUMAR; BHAN, 2018).

Uma das vantagens em usar três convoluções  $3 \times 3$  em vez de uma convolução  $7 \times 7$  é que aplicando três camadas ReLU ao invés de uma, torna-se mais discriminativa a função de decisão. Outro ponto é que reduz os parâmetros (YANG et al., 2018; DHOMNE; KUMAR; BHAN, 2018).

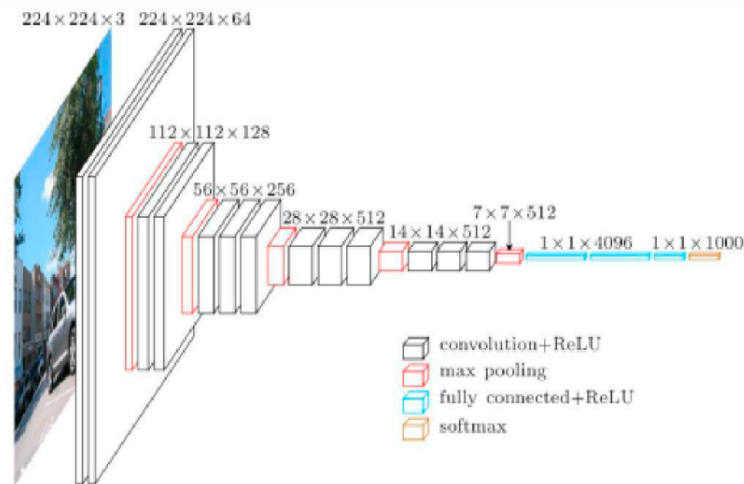


Figura 10 – Arquitetura da rede VGGNet. Fonte: (DHOMNE; KUMAR; BHAN, 2018).

#### 2.3.2.4.6 ResNet

A ResNet ou *Deep Residual Network* foi desenvolvida com o intuito de ser um sistema Neural Artificial com camadas mais profundas de alta precisão (Budhiman; Suyanto; Arifianto, 2019). A ideia da ResNet é elaborar um sistema artificial em que se possa atualizar seu peso nas camadas mais rasas, reduzindo-se o chamado "gradiente de desaparecimento", problema no qual redes tradicionais muito profundas convergem adicionando resíduos no treinamento, ou seja, ligações que transitam pela rede sem passar por funções não-lineares (He et al., 2016). Este conceito é implementado usando um "atalho conexão", como observado na Figura 11 (Budhiman; Suyanto; Arifianto, 2019). Esse conjunto é composto por blocos de construção, duas camadas não-lineares em que se passa uma dada entrada  $X$ , definidos como  $F(x)$ . Na saída, os dados são somados com os valores da própria entrada  $X$ , designado como residual.

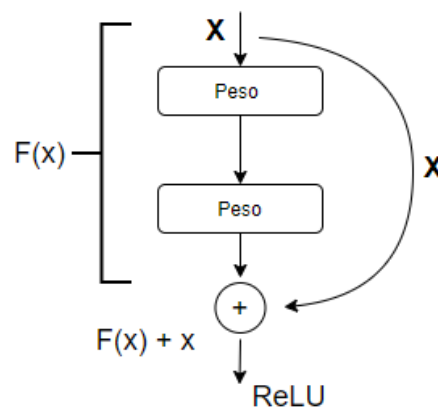


Figura 11 – Bloco de construção ResNet. Fonte: Traduzido de (Budhiman; Suyanto; Arifianto, 2019).

A implementação dessa arquitetura constitui de uma camada de entrada por uma

convolução, a fim de diminuir a complexidade, seguido por vários blocos de construção. No final, o sinal de saída é entregue pela classificação dos dados de entrada, realizado por uma camada totalmente conectada (MARTINS, 2018).

#### 2.3.2.4.7 FayNet

A FayNet é uma CNN desenvolvida por Rodrigo Fay Vergara em 2018, baseada na LeNet de LeCun. A escolha desta arquitetura, segundo (VERGARA, 2018), foi devido ao fato de possuir um banco de dados limitado com poucas imagens para treinamento da rede, e pela mesma ser de fácil compreensão. Outro motivo é que foi possível construir uma rede não muito profunda como desenvolvido em outros trabalhos de aprendizado profundo, evitando o *overfitting*.

O *overfitting* é um dos incômodos mais persistentes no aprendizado de máquina e acontece quando o modelo utilizado não generaliza bem os dados de treinamento para dados não vistos, ou seja, no modelo treinado ele é preciso, porém, quanto adicionado dados não vistos ele apresenta uma boa classificação.

O modelo desenvolvido é formado por 4 camadas de aprendizado com pesos, sendo 2 camadas de convolução seguidas por funções não-lineares, e duas camadas de conexão total. O resultado da última camada de conexão total alimenta a camada de saída com uma função de *softmax* de 2 variáveis que serão identificadas, portadores ou não de esquizofrenia. Na Figura 12 está ilustrada a arquitetura da FayNet.

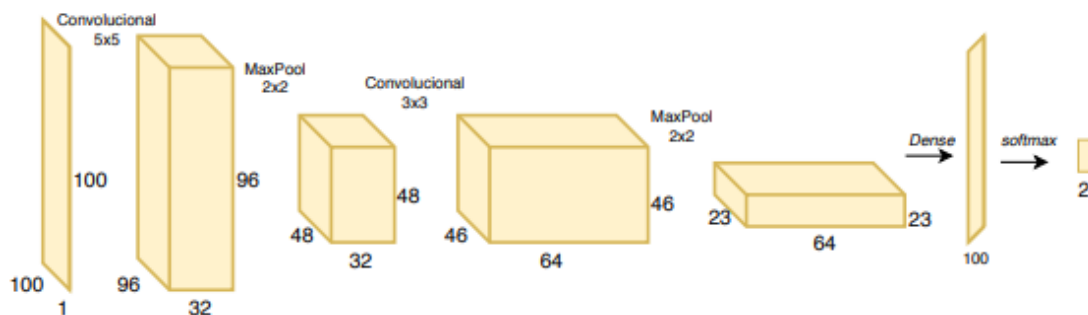


Figura 12 – Arquitetura da FayNet. Fonte: (VERGARA, 2018)

Todas as imagens do banco de dados foram redimensionadas para o tamanho 100x100, com o intuito de diminuir a complexidade e a demora no treinamento da rede. O tamanho das imagens para a entrada do sistema é de 100x100, mas uma vez que passam pela camada de convolução com 32 filtros de tamanho 5x5, com passo de 1 e *zero-padding* de 0, o tamanho é reduzido para 96x96 para que seu arranjo espacial seja mantido, eliminando pixels das bordas. Nota-se que as MRI utilizadas possuem bordas pretas irrelevantes ao problema, este 46 corte não influencia no resultado, mas sim diminui os parâmetros de aprendizado, reduzindo o *overfitting* (VERGARA, 2018).

## 2.4 Treinamento

O treinamento de um modelo CNN é realizado por um algoritmo chamado de retro-propagação, ou do inglês *back propagation*. Esse treinamento é realizado em um determinado conjunto de dados composto por imagens e rótulos, como classes e máscaras. Ao iniciar o treinamento do modelo, os coeficientes das camadas são inicializados aleatoriamente. O algoritmo calcula um valor de erro na saída da última camada para atualizar os pesos de cada neurônio desta camada. Com esses pesos atualizados, é calculado um novo valor de erro e os pesos da camada anterior são atualizados. Esse procedimento é realizado até que o cálculo dos pesos sejam propagados até a primeira camada (WITTEN et al., 2017).

Em algoritmos de aprendizado de máquina, o objetivo é minimizar a função de perda para diminuir a probabilidade de erro. A função de perda é um método para calcular o erro da previsão da rede. A função mais adequada para problemas de classificação binária é a perda de entropia binária cruzada, ou *Binary Cross-Entropy BCE*, uma representação da *Cross-entropy CE* ou entropia cruzada com duas classes, dada pela equação 2.6 (LIN et al., 2018)

$$CE(p, s) = - \sum_i^C p_i \log(s_i) \quad i \in [1, N], \quad (2.6)$$

em que  $p_i$  é o valor real da classe comparada ou verdade fundamental,  $S_i$  é a predição referente a classe através da descrição de probabilidade de função de ativação, como por exemplo a *softmax* e *sigmoid*, e  $C$  o número de classes. Para uma abordagem binária,  $C = 2$ , temos a equação 2.7.

$$CE = -p_1 \log(s_1) - p_2 \log(s_2), \quad (2.7)$$

onde a soma da probabilidade de ocorrer as duas classes  $[0,1]$  é igual a 1, ou seja 2.8,

$$p_1 + p_2 = 1. \quad (2.8)$$

Logo, ao aplicar em 2.6 chega-se à equação BCE 2.9

$$CE(p, s) = - \sum_i^{C=2} p_i \log(s_i) = -p_1 \log(1 - s_1) - (1 - p_1) \log(1 - s_1), \quad (2.9)$$

$$BCE = -\log(s_1) \quad \text{se } p = 1, \quad (2.10)$$

$$BCE = -\log(1 - s_1) \quad \text{se } p = 0. \quad (2.11)$$

O bom resultado do treinamento de um modelo também se dá pelos ajustes de outros hiper-parâmetros, como os otimizadores, taxa de aprendizagem e tamanho do lote, ou *batchsize*, que é o número de exemplos de treinamento usados a cada iteração.



Otimizadores são ferramentas usadas para adequar os atributos da rede neural, como os pesos e taxa de aprendizagem, com o objetivo de reduzir as perdas. Existem variados tipos de otimizadores como os comumente utilizados Descida do Gradiente Estocástico (ou do inglês SGD, *Stochastic gradient descent*) e o *RMSprop*. A taxa de aprendizagem é uma variável na qual é definido o passo para alterar o valor dos pesos. Se esse passo for muito grande, há a probabilidade da rede ignorar um valor ideal para um determinado peso, o que afeta de maneira direta a capacidade de minimizar a função de perda. Esse valor é um número quase sempre pequeno. Seu uso mais popular é um valor em torno de 0,001 (Cheng Lv; YI; TAN, 2006).

## 2.5 Ferramentas de Validação do modelo

### 2.5.1 Validação cruzada K-fold

A validação cruzada *k-fold* é uma ferramenta muito utilizada na avaliação de modelos de aprendizado de máquina para um conjunto de dados limitado. O método consiste no particionamento do conjunto de dados, de forma embaralhada e que não haja sobreposição de dois conjuntos de teste, em *k* subconjuntos distintos com tamanhos iguais ou quase iguais chamados *folds* (BERRAR, 2018). Um exemplo desta divisão de dados pode ser observada na Figura 14.

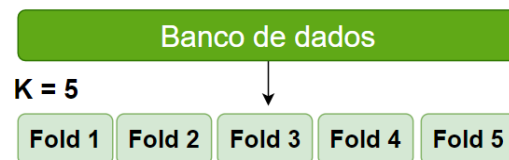


Figura 13 – Exemplo de partição na validação cruzada com  $k=5$ . Fonte: Autor

Com as divisões, o treinamento e teste são realizados em *k* interações, onde em uma, é separada um *fold* para testar o modelo, e as outras *k-1* partições restantes destinadas ao treinamento. As métricas como acurácia e precisão são calculadas em cada interação, e a média desses valores validará o desempenho do modelo (YADAV; SHUKLA, 2016).

### 2.5.2 Métricas de desempenho

O êxito de uma classificação é dado através da sua avaliação, sendo ela realizada por métricas de desempenho. Uma das principais métricas para qualificar classificadores de aprendizagem supervisionada é a matriz de confusão, uma representação na forma de tabela do desempenho de um algoritmo de classificação, de modo a compreender a relação entre os erros e acertos do modelo.



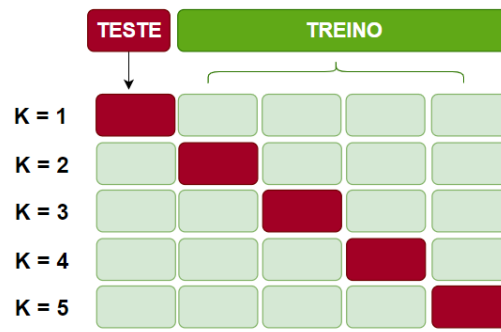


Figura 14 – Exemplo da separação de *folds* para treinamento e validação de um modelo na validação cruzada com  $k = 5$ . Fonte: Autor

Para um melhor entendimento da matriz é necessário compreender o conceito de positivo e negativo neste ambiente. A classe de interesse no processo de classificação é tratado como positivos, neste caso pacientes com esquizofrenia, já os negativos são designados a outra classe, pacientes ditos como saudáveis. A matriz de desempenho é desenvolvida com os valores organizados em quatro grupos, sendo:

- **Positivo verdadeiro (True positive TP)** - O dado foi dito como verdadeiro assim como previsto;
- **Falso positivo (False positive FP)** - O dado era negativo porém foi previsto como positivo;
- **Negativo verdadeiro (True negative TN)** - O dado foi dito como negativo assim como previsto;
- **Falso negativo (False negative FN)** - O dado era positivo porém foi previsto como negativo.

Tabela 1 – Matriz de confusão

Rótulo	Previsão	
	Positivo	Negativo
Positivo	TP	FN
Negativo	FP	TN

A partir da matriz de confusão, exibida na Tabela 1, é possível abstrair diversas informações como as métricas acurácia, sensibilidade, precisão e especificidade. A acurácia ( $A$ ) contabiliza a quantidade de acertos em relação ao total de classificações, e é dada por

$$A = \frac{TN + TP}{TN + TP + FN + FP}. \quad (2.12)$$

A sensibilidade ( $S$ ) analisa a qualidade do classificador para classificar o grupo de interesse, ou seja, é a porção de TP em relação ao total de positivos, representada por

$$S = \frac{TP}{TP + FN}. \quad (2.13)$$

A precisão ( $P$ ) é a fração de TP em relação aos classificados como positivos, em outras palavras é a quantidade positiva classificada corretamente e é descrita como

$$P = \frac{TP}{TP + FP}. \quad (2.14)$$

A especificidade ( $E$ ) apresenta uma análise contrária à sensibilidade, tem o objetivo de observar o quão bom o classificador é para classificar corretamente a outra classe. É calculado como a porção de TN em relação ao total de negativos, ou seja

$$E = \frac{TN}{TN + FP}. \quad (2.15)$$

Uma outra métrica interessante de se observar é a métrica *F1-Score*, ela busca avaliar o equilíbrio da rede em relação a sua precisão e sensibilidade. Essa relação é dada pela equação 2.16.

$$F1 = \frac{2 \times P \times S}{P + S}. \quad (2.16)$$

A medida *F2-score* balanceada é um método que combina precisão e sensibilidade. A pontuação F2 calcula uma média harmônica ponderada entre essas métricas (DEVARRIYA et al., 2020). Nela é atribuída maior relevância à sensibilidade do que à precisão. O funcionamento desse modelo de avaliação é dado pela equação 2.17

$$F2 = \frac{1 + \beta^2 \times P \times S}{\beta^2 \times P + S}, \quad (2.17)$$

onde,  $\beta$  é uma constante que controla a ponderação, geralmente é utilizado o valor de  $\beta = 2$ .

### 2.5.3 Curva Característica de Operação do Receptor (ROC) e Área Sob a Curva (AUC)

A curva característica de operação do receptor (do inglês ROC, *Receiver Operating Characteristic*), é um gráfico que tem como objetivo ajudar a analisar o desempenho de um classificador, geralmente binário. Ele traça uma curva de probabilidade contendo a taxa de verdadeiros positivos (TVP) versus a taxa de falsos positivos (TFP) em vários valores de limiar. A Figura 15 demonstra a curva que representa essa relação.

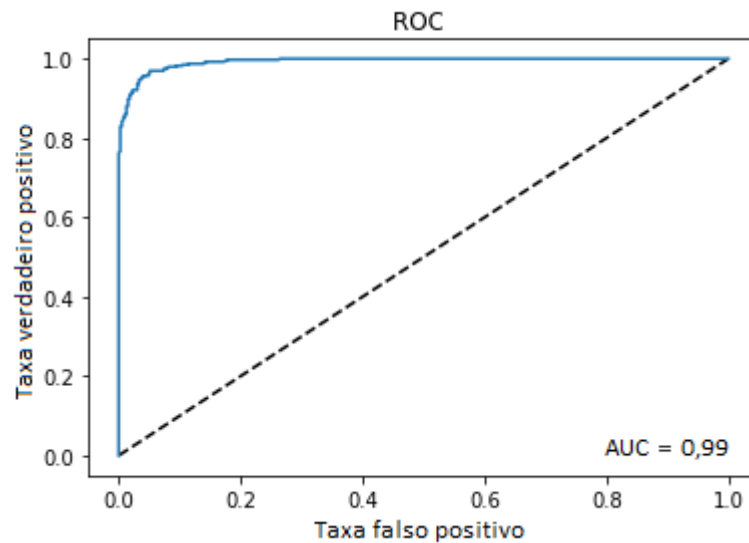


Figura 15 – Exemplo ROC e área sobre a curva AUC próxima à ideal. Fonte: Autor

A área sob a curva (do inglês, *Area under the Curve* - AUC) é a medida da capacidade de um classificador de distinguir entre as classes, e é usada como um resumo da curva ROC. O valor da AUC é um valor entre 0 e 1, de forma que quando a AUC é ideal, ou seja igual a 1, o classificador consegue prever corretamente durante o teste quais dados são verdadeiros positivos sem a existência de falsos positivos.



## 3 Materiais e Métodos

Este capítulo apresenta os procedimentos e o desenvolvimento dos sistemas de testes dos métodos de classificação automática, com os bancos de dados disponibilizados através do *SchizConnect*, um banco de dados virtual que contém neuroimagens de pacientes portadores de esquizofrenia, exibindo os primeiros ensaios e os detalhes das redes neurais utilizadas.

Aqui também é descrito a estratégia detalhada para a realização dos testes, como o pré-processamento e a organização das imagens, além da exibição do banco de dados disponibilizado, como a quantidade de imagens, quantidade de indivíduos diagnosticados com a doença (esquizofrenia) e indivíduos de controle (saudáveis), compreendendo a idade e o gênero dos pacientes.

### 3.1 Banco de dados

O banco de dados utilizado para o desenvolvimento deste trabalho é do *SchizConnect* (WANG et al., 2018), um banco de dados virtual de pesquisa e *download* para dados públicos de neuroimagem de esquizofrenia. Esta plataforma utiliza um *software* que realiza a integração de dados relacionados a esquizofrenia a partir de instituições parceiras como: fBIRN, COINS, XNAT Central, NUNDA e NU REDCap.

O *SchizConnect* traz a possibilidade da exploração e análise em grande escala de dados multi-locais, multi-dimensionais e multimodais, de uma maneira mais simples para a compreensão de pesquisadores e profissionais de saúde, que estudam e desenvolvem trabalhos sobre esta doença complexa.

A organização de dados dentro da plataforma está ilustrado na Figura 16 e na Tabela 2, onde os dados são divididos pelas variações da esquizofrenia e outros distúrbios, além de detalhar a faixa etária e o gênero dos pacientes envolvidos. As variações do banco de dados são dados por: Nenhum transtorno conhecido (NTC), esquizofrenia ampla (ESQA), esquizofrenia rigorosa (ESQR), esquizoafetivo (ESQ), transtorno bipolar (TP), irmão da esquizofrenia ampla (IEA), irmãos da esquizofrenia estrita (IEE) e irmão de nenhum transtorno conhecido (INTC).

Os arquivos são disponibilizados no formato *NII*, um formato de arquivo aberto associado ao *Neuroimaging Informatics Technology Initiative (NIFTI-1)* (NIFTI, 2020), utilizado para armazenar dados de imagens cerebrais obtidos por ressonância magnética. Esses arquivos possuem uma representação completa em 2D e 3D do cérebro em três eixos, representando a parte lateral, frontal e superior do cérebro como apresentado na

Figura 17, onde é possível observar as imagens por fatias em 2D através do *software* de visualização *online Statistics Online Computational Resource (SOCR)*.

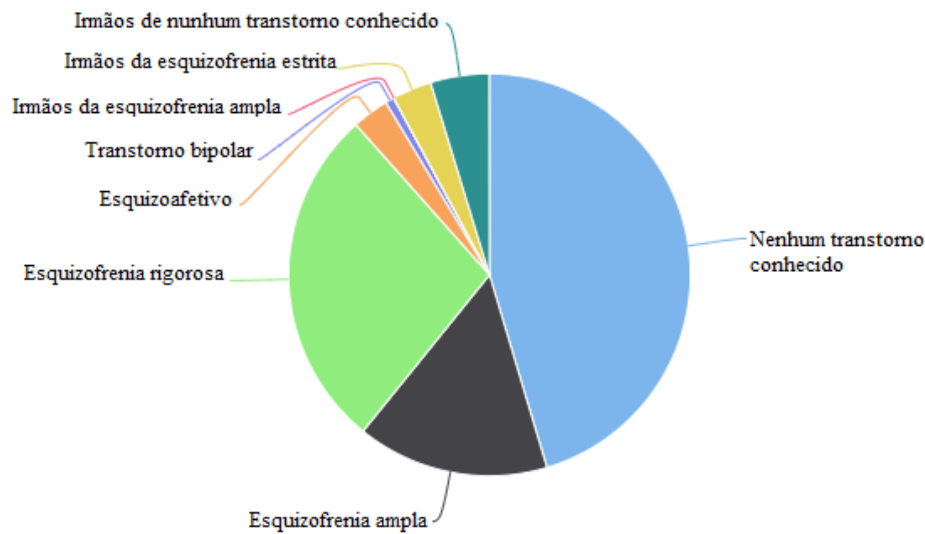


Figura 16 – Organização dos dados no SchizConnect. Fonte: Traduzido de (WANG et al., 2018)

Tabela 2 – Organização detalhada dos dados no SchizConnect . Fonte: Adaptado de (WANG et al., 2018)

Estatística / DX	NTC	ESQA	ESQR	ESQ	TP	IEA	IEE	INTC
Tamanho dos dados	632	215	384	41	10	0	44	66
Sexo (m/f)	392/240	173/42	278/106	25/15	5/5	0/0	21/23	16/15
Idade (anos) (média $\pm dp$ )	34,7 $\pm$ 12,6	34,9 $\pm$ 12,9	35,3 $\pm$ 12,4	39,2 $\pm$ 10,0	46,6 $\pm$ 13,4	-	21,6 $\pm$ 3,7	20,4 $\pm$ 3,5
Faixa etária (min-max)	13 - 67	0 - 65	17 - 66	19 - 59	21 - 64	-	14 - 28	14 - 28

Como o conjunto de dados é formado por instituições diferentes, as imagens possuem a quantidade de cortes ou fatias distintas nos três eixos, porém com a mesma resolução 256x256 e ponderadas em T1 e T2. Para o COBRE, o número de fatias por eixo (Coronal, Sagital e Transversal) é de 246, 268 e 223 respectivamente. No FBIRN é de 145, 238 e 240. Já no N MORPH, banco de dados do NUNDA e NU REDCap, é 264, 260 e 197, e no NUSDAST, banco de dados do instituto XNAT Central, é de 180 nos três eixos. Para o desenvolvimento deste trabalho, foram utilizadas 968 imagens de ressonância magnética em T1, sendo 484 de pacientes esquizofrênicos e 484 pacientes sem nenhum transtorno.

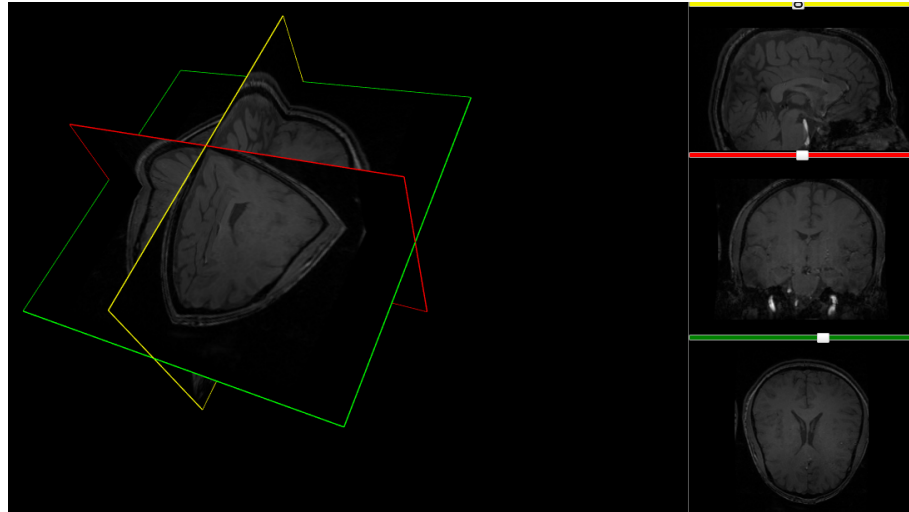


Figura 17 – Visualização 3D a partir do software 3D *SOCR* de uma imagem de ressonância magnética do banco de dados do *SchizConnect*. Fonte: Autor.

## 3.2 Pré-processamento

As etapas de pré-processamento são extremamente necessárias para adequar as imagens aos formatos compatíveis com o que as redes neurais estão preparadas para receber. O banco de dados apresentado é formado por neuroimagens em 3D, onde possuem informações ao longo dos 3 eixos, o Coronal, Sagital e Transversal. Porém, como as redes a serem utilizadas possuem arquitetura 2D, ou seja, esperam como entrada imagens com informações em apenas duas dimensões é necessário realizar o corte dos dados tridimensionais em fatias, adequando-os assim ao formato desejado.

Os cortes escolhidos não foram sequenciais já que cada neuroimagem é composta por mais de 100 fatias ao longo de cada eixo, e a distância entre uma fatia da outra é muito pequena. Assim, os cortes próximos uns dos outros, como o 15 e 16 por exemplo, apresentam dados muitos semelhantes. Devido a isso serão utilizados os cortes 15 e 20 no eixo Transversal, por evidenciarem o ventrículo cerebral e possuírem características diferentes entre si. Já a utilização de mais de um corte que evidencia o ventrículo, faz com que a rede tenha mais dados para aprender as características necessárias para realizar a classificação.

O *software* Clínica será utilizado para essa etapa de pré-processamento, uma vez que já possui diversos algoritmos de processamento já implementados e um sistema compatível com o padrão *Brain Imaging Data Structure* - BIDS (GORGOLEWSKI et al., 2016) de organização de imagens, que facilita substancialmente o processamento e gerenciamento dos dados. No Clínica, será aplicado o *pipeline t1-linear*, onde são executadas etapas para alinhar imagens de ressonância magnética ponderadas em T1 ao modelo MNI (MNI152NLin2009cSym), um gabarito contendo as coordenadas e regiões cerebrais como um quadro de referência artificial, desenvolvido para descrever diferentes anatomias de

forma unificada (FONOV et al., 2009). Este alinhamento é realizado através da biblioteca ANTs, uma ferramenta de normalização para o gerenciamento e visualização de dados multidimensionais, como as neuroimagens (AVANTS et al., 2014). Outra ferramenta oferecida pelo *software* é a normalização espacial das imagens através do mapeamento paramétrico estatístico (do inglês SPM, *Statistical Parametric Mapping*).

A saída do pré-processamento no Clínica são neuroimagens multidimensionais, ainda em 3D, alinhadas e padronizadas, de modo com que todos cortes representam as mesmas regiões em todas as neuroimagens. Para a captura das imagens em 2D dos cortes a serem trabalhados, utilizou-se a biblioteca de documentação aberta *Nilearn*, uma metodologia estatística e de aprendizado de máquina que permite análises acessíveis e versáteis de volumes cerebrais (ABRAHAM et al., 2014). Essa etapa de pré-processamento garante que a região de interesse, no caso o ventrículo, seja a mesma em todas as imagens. O conjunto de imagens serão salvos no formato *png* com a dimensão de  $229 \times 220$  *pixels*, em um total de 1936 imagens.

Além de adequar o formato das imagens é necessário também na etapa de pré-processamento realizar a organização dos dados de forma que a rede possa entender quais as imagens que possuem rótulos referentes a pacientes esquizofrênicos e quais são do grupo de controle saudável.

### 3.3 Redes a serem testadas

A fim de realizar a classificação de imagens para auxílio no diagnóstico de esquizofrenia, foram pesquisados trabalhos que utilizaram redes neurais convolucionais em processamentos semelhantes ao que buscamos, trabalhando com a classificação a partir de imagens de ressonância magnética cerebral. Dentre as diversos trabalhos analisados, os que apresentaram melhor descrição das etapas de pré processamento de dados e de treinamento, com uma boa acurácia na classificação, foram os trabalhos de Vergara (VERGARA, 2018) e Fung (FUNG et al., 2019). Mais especificamente, o trabalho de Vergara apresentou a maior quantidade de parâmetros utilizados para sua arquitetura, o que possibilita sua reconstrução.

A utilização de outras redes, como algumas de arquitetura 3D, também foram consideradas. No entanto, a relação entre custo computacional e melhoria da precisão de classificação era baixa quando comparada a arquiteturas 2D, cerca de 1%, como mostra o trabalho de (FUNG et al., 2019). Além disso, as máquinas que temos a disposição para realizar a classificação não possuem poder de processamento necessário para viabilizar a utilização das arquiteturas 3D.

Alguns serviços de processamento em nuvem, como o *Google COLAB*, foi cogitado para realização da implementação de arquiteturas 3D, no entanto é uma ferramenta que



possui limite de utilização em sua versão gratuita.

### 3.3.1 FayNet

A FayNet (VERGARA, 2018) foi desenvolvida para a classificação de pacientes com esquizofrenia e pacientes saudáveis através de MRI cerebral, com resultado de 77,85 % de acurácia para a classificação. O sistema foi testado com o banco de dados disponibilizado da *Function Biomedical Informatics Research Network* (FBIRN) (GLOVER et al., 2012) contendo 153 imagens de controle e 160 imagens de pacientes esquizofrênicos. A partir do banco de dados, organizou-se as imagens de entrada para se adequar a arquitetura, em que só foram utilizados os cortes axiais 13 e 14 por apresentarem de maneira mais clara o ventrículo do cérebro, a região responsável por produzir maiores resultados relevantes ao assunto, visto na literatura.

Para o treinamento e a validação do sistema, foram realizadas duas análises. Na primeira análise utilizou-se a técnica de *holdout*, onde foram criados dois vetores para a rotulação das classes, sendo 1 para os pacientes esquizofrênicos e 0 para a classe de pacientes saudáveis, com uma separação de dados de 70% das imagens para treinamento e 30% para realização dos testes. Com o vetor rotulado, seus valores foram normalizados, subtraindo a média das imagens indicadas pelo vetor e dividindo pelo desvio padrão, obtendo imagens entre -1 e 1, facilitando o treinamento da rede por diminuir a variação destes valores. Já para o segundo estudo foi empregado o método de validação cruzada *k-fold* com  $k = 7$ , para dar confiabilidade no desempenho da rede.

Os algoritmos foram desenvolvidos em Python usando o compilador Theano, e os pacotes Lasagne (DIELEMAN et al., 2015) e nilearn (NOURI, 2014) para o desenvolvimento e treinamento da rede neural e extração das características das camadas internas da rede. O Lasagne é um *framework* específico para treinamento de redes neurais que utiliza a biblioteca de funções matemáticas Theano (TEAM, 2016).

No entanto, como o Theano está, nos dias de hoje, ultrapassado e em desuso, neste trabalho a FayNet será implementada no *framework* Tensorflow e no Keras, uma API de alto nível do TensorFlow para criar e treinar modelos de aprendizado profundo, por possuir uma abordagem de prototipagem rápida, facilidade de pesquisa e produção (ABADI et al., 2015). A rede foi totalmente implementada a partir das descrições de arquitetura (VERGARA, 2018), além de propormos algumas modificações em suas configurações, como na camada de *pooling* e camada densa, com o objetivo de melhorar o desempenho.

Logo após as camadas convolucionais, foram adicionadas camadas de normalização em lote (ou do inglês, *batch normalization*), que busca otimizar o treinamento da rede com o equilíbrio no processo de aprendizagem, no intuito de reduzir o número de períodos nesta etapa. Já na camada de *pooling*, para a redução de dimensionalidade das imagens,

a operação *MaxPooling* foi modificada para a *AveragePooling*, onde o cálculo é realizado pela média dos valores de cada *path* no mapa de recursos.

Na camada densa a proposta está na troca da função de ativação. Na FayNet (VERGARA, 2018) é utilizada a função *softmax*, bastante utilizada para problemas de classificações com múltiplas classes. Já a função de ativação utilizada neste novo modelo será a função *sigmoid*, que segundo (ABADI et al., 2015), é a mais adequada para a nossa abordagem, onde o objetivo é realizar uma classificação binária.

Para o treinamento desta nova versão, utilizou-se o mesmo otimizador da rede FayNet, gradiente descendente estocástico (do inglês SGD, *Stochastic Gradient Descent*), com a mesma taxa de aprendizagem  $lr = 0,014$ .

Na Figura 18 é apresentado o diagrama do modelo, assim como a proposta de melhoria, no qual observa-se uma comparação nítida de todas as suas modificações. Onde as camadas em vermelho, foram as camadas adicionadas e as camadas em verde, foram as camadas que tiveram a função de *pooling* modificadas. Com o intuito de comparar as duas redes implementadas com Tensorflow, ambas as versões foram avaliadas utilizando os mesmos bancos de dados, tanto o FBIRN quanto o *SchizConnect*, além de treinadas utilizando 150 épocas e *batchsize* igual a 90.

### 3.3.2 ResNet18

A utilização dessa arquitetura se dá devido ao fato dela ter apresentado resultados satisfatórios em uma tarefa de classificação de imagens similar a que desejamos realizar, como por exemplo no trabalho de (FUNG et al., 2019) que buscou classificar MRI cerebral. Esse estudo tinha como objetivo classificar a doença de Alzheimer em três categorias, alcançando cerca de 84% de acurácia.

A rede foi implementada a partir do modelo da ResNet18 convencional (JIANG, 2020), utilizando a biblioteca de código aberto *Tensorflow* com o *Keras*, que facilita a construção de algoritmos de aprendizado de máquinas utilizando linguagem *Python*. A arquitetura do modelo a ser implementada e testada está descrito no diagrama de blocos da Figura 19, onde estão explicitados os parâmetros de toda a arquitetura, bem como os *kernels* e seus respectivos tamanhos.

Além do modelo descrito na Figura 19, outros parâmetros relevantes utilizados para o treinamento da arquitetura são: a função de otimização e a função de classificação binária. O otimizador a ser utilizado é o otimizador gradiente descendente estocástico, com uma taxa de aprendizado  $lr = 0.001$ . Quanto a função de interpretação de distribuição de probabilidade utilizada, neste caso binária, será a *sigmoid*.

Para o treinamento da rede também será utilizado o banco de dados *SchizConnect*, com 120 épocas e *batchsize* igual a 40. A performance do modelo será avaliada pela

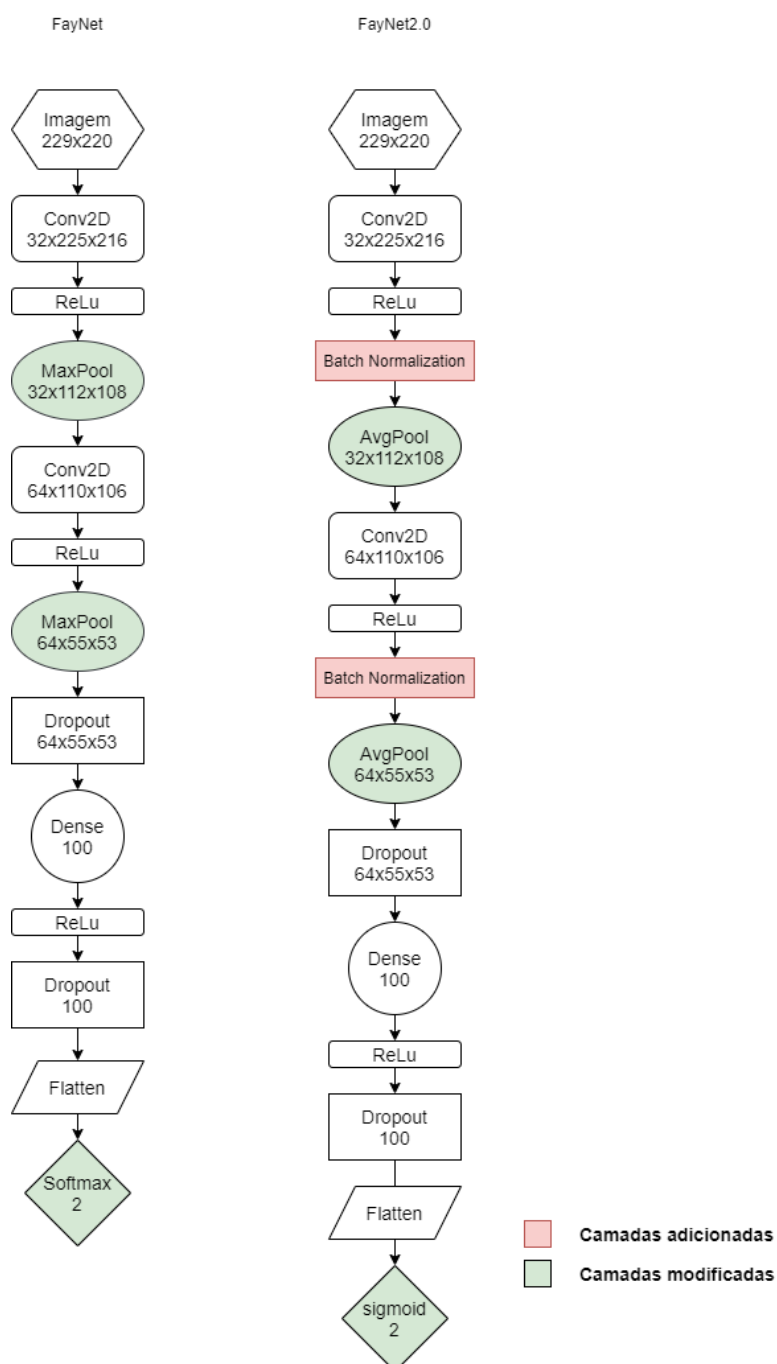


Figura 18 – Detalhes do modelo Faynet vs Faynet2.0. Fonte: Autor.

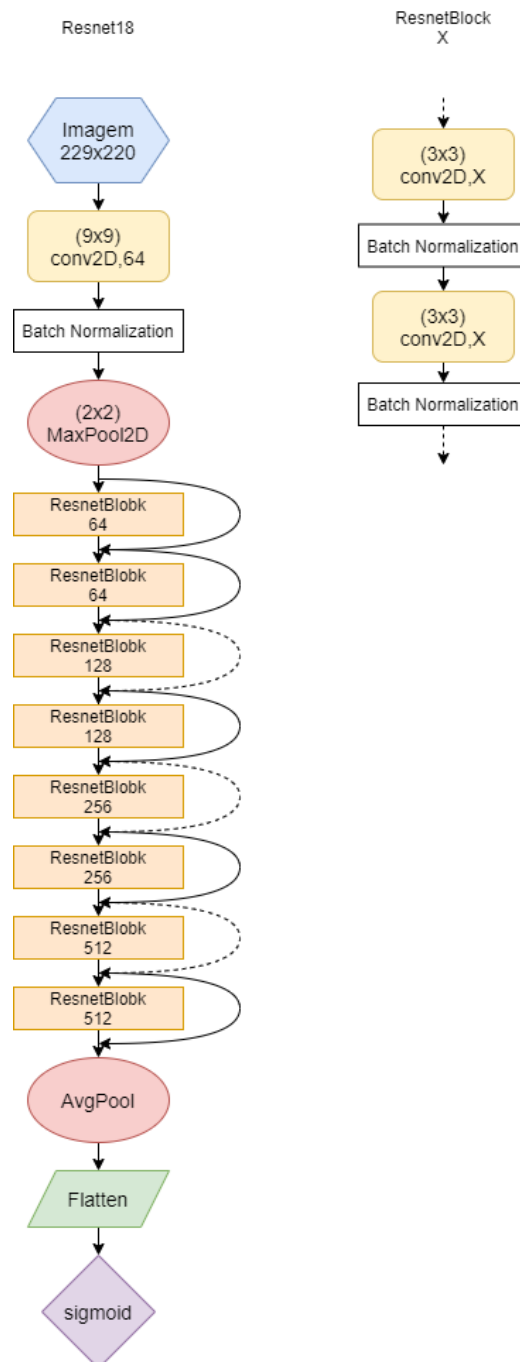


Figura 19 – Diagrama contendo modelo da ResNet18 com os *kernels* utilizados para convolução e *pooling*. Fonte: Autor.

validação cruzada com  $k = 7$ , onde retirando-se o *fold* destinado ao teste serão utilizadas 20% das imagens restantes para a validação do modelo.

## 3.4 Avaliação de desempenho

Para a validação dos modelos apresentados neste documento foi utilizado um *k-fold* com  $k = 7$ , onde foram embaralhados e separados de forma aleatório para realizar o treinamento e o teste das redes aqui descritas. Assim com os dados do *SchizConnect*, das 1936 imagens disponíveis, resultarão em 7 conjuntos de 276 imagens, onde a cada *fold* de teste, serão calculadas as métricas de desempenho resultantes do treinamento com os outros 6 *folds*.

A partir da matriz de confusão obtida pelo treinamento da rede com os 7 *folds*, e utilizando as equações 2.12, 2.14, 2.13, 2.15 e 2.16 as métricas de acurácia, precisão, sensibilidade, especificidade e a f2-score poderão ser calculadas, assim como suas médias e desvio padrão. Com essas métricas será possível avaliar os modelos de CNN's. Além dessas métricas a ROC e a AUC serão calculadas, para uma representação visual dos parâmetros da rede. Essas métricas foram escolhidas com o intuito de melhorar a avaliação e aumentar a credibilidade dos resultados obtidos com os testes de classificação.



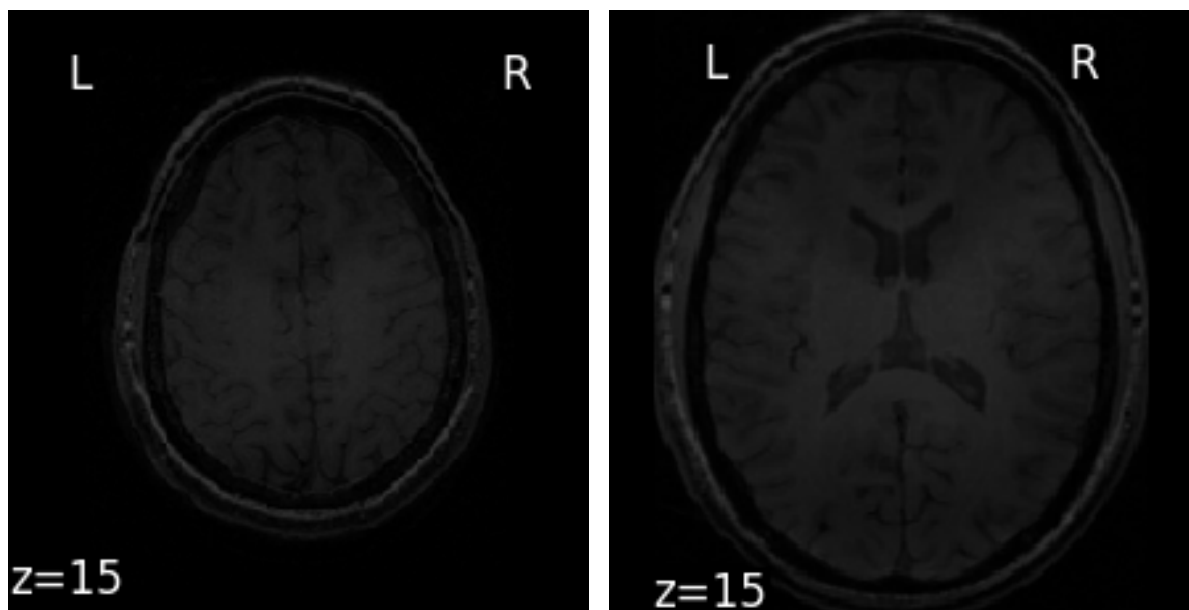
## 4 Resultados e Discussões

Neste capítulo serão apresentados os resultados e análises da classificação de pacientes esquizofrênicos e saudáveis através das redes propostas, assim como as etapas de pré-processamento, sendo eles observados em diagramas, curvas de aprendizado, perdas e métricas de desempenho. Os resultados de cada rede serão comparados entre si por ferramentas de validação, como a validação cruzada *K-fold*.

### 4.1 Pré-processamento

Como já mencionado, as entradas das redes propostas são imagens em 2D, porém o banco de dados utilizado é composto por neuroimagens tridimensionais. Para a seleção das fatias que evidenciam o ventrículo cerebral foi necessário o pré-processamento de um *pipeline* no *software* Clínica que realiza a padronização da imagens, conforme descrito na seção 3.2.

Para uma melhor compreensão desta etapa de pré-processamento, nas Figuras 20 e 21 estão expostas duas imagens do banco de dados no qual pode-se comparar o antes e depois do pré-processamento. Para todos os casos foram escolhidos o eixo transversal e corte 15, representado por  $z = 15$ .



(a) Corte  $z=15$  antes do pré-processamento.

(b) Corte  $z=15$  depois do pré-processamento.

Figura 20 – Etapa de pré-processamento da imagem 1. Fonte: Autor.

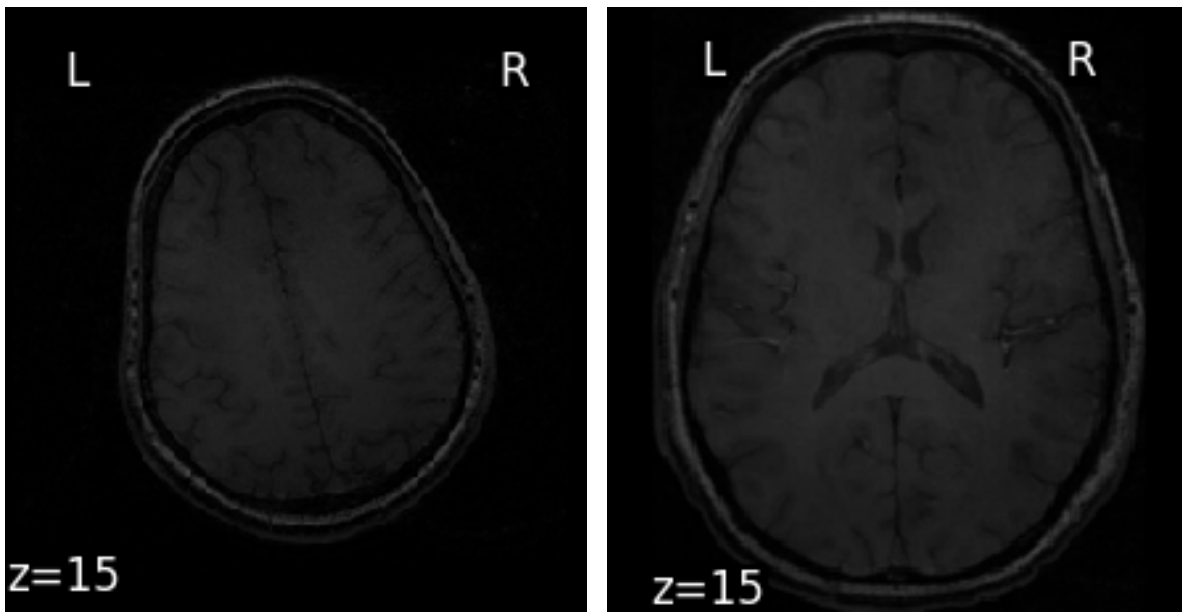
(a) Corte  $z=15$  antes do pré-processamento.(b) Corte  $z=15$  depois do pré-processamento

Figura 21 – Etapa de pré-processamento da imagem 2. Fonte: Autor.

As Figuras 20a e 21a são referentes as imagens antes do processo de *pipeline* do *software* Clínica, no qual observa-se, em ambos os casos, que as imagens no corte estabelecido não apresentam o ventrículo cerebral, região de análise para o diagnóstico proposto. Já nas Figuras 20b e 21b pode-se notar o resultado do procedimento, em que no mesmo corte ou fatia, é apresentada a mesma região morfológica cerebral, exibindo-se a região de interesse.

## 4.2 FayNet

Através dos procedimentos relatados na subseção 3.3.1 os modelos FayNet foram implementados e comparados com as suas respectivas características. Para uma melhor identificação, a nova versão proposta neste trabalho é chamada de FayNet 2.0. Como o código fonte da FayNet não foi utilizado por conta do uso de um *framework* ultrapassado, a rede foi implementada de forma idêntica às descrições de sua arquitetura e treinamento (VERGARA, 2018).

Para a validação dos modelos, foram realizadas alguns testes comparando seus desempenhos com dois banco de dados distintos. Primeiro os modelos foram implementados utilizando o mesmo banco de dados, com os mesmos cortes, usado para a validação da rede original (VERGARA, 2018), o FBIRN. O banco de dados utilizado é composto por 312 imagens, sendo 152 pacientes portadores de esquizofrenia e 160 pacientes de controle. As imagens são compostas por 27 cortes axiais e para este trabalho foram usados os mesmos cortes, 13 e 14 utilizados por Vergara.



Tabela 3 – Arquitetura da rede FayNet com os parâmetros para o banco de dados FBIRN.

Nome (Camada)	Tamanho filtro/passo	Tamanho de saída	Parâmetros de aprendizado
Entrada		1X256X256	0
Conv2d	5X5/1	32X252X252	832
max_pooling2d	2X2/1	32X126X126	0
Conv2d	3X3/1	64X124X124	18496
max_pooling2d	2X2/1	64X62X62	0
dropout1			0
dense			6500
dropout			0
flatten			0
Saída			768802
<b>Total</b>			<b>794630</b>

As arquiteturas das redes executadas podem ser observadas nas Tabelas 3 (FayNet) e 4 (FayNet 2.0), onde são detalhadas as principais características dos sistemas para cada camada, como tamanho dos filtros e passos, a dimensão de saída e quantidade de parâmetros de aprendizado.

Na Tabela 5 são apresentadas as métricas de desempenho dos modelos para este teste, como acurácia (AC), precisão (PR), especificidade (ESP), sensibilidade (SEN),  $F1$  - score (F1) e  $F2$  - score (F2) com o respectivo desvio padrão. Essas métricas foram calculadas com a média dos resultados obtidos a partir do método de validação cruzada  $k$ -fold com  $k=7$ .

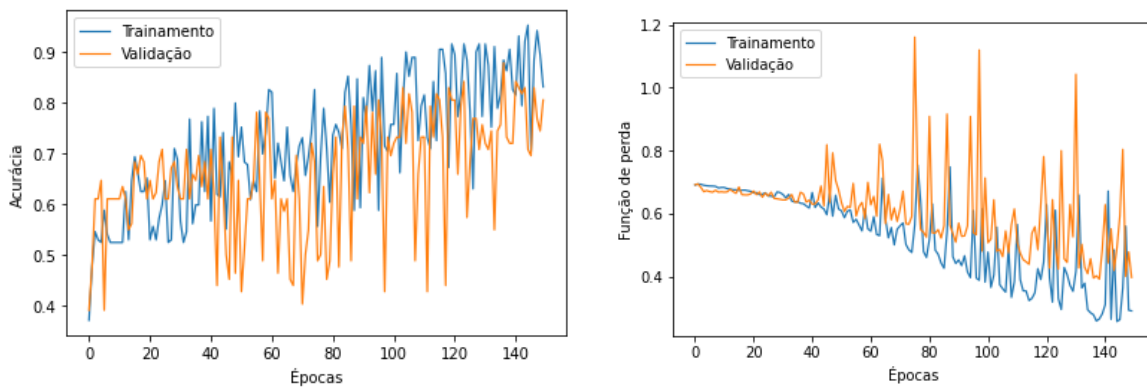
Tabela 4 – Arquitetura da rede FayNet 2.0 com os parâmetros para o banco de dados FBIRN.

Nome (Camada)	Tamanho filtro/passo	Tamanho de saída	Parâmetros de aprendizado
Entrada		1X256x256	0
Conv2d	5X5/1	32X252X252	832
batch_normalization			128
average_pooling2d	2X2/1	32X126X126	0
Conv2d	3X3/1	64X124X124	18496
bach_normalization			256
averege_pooling2d	2X2/1	64X62X62	0
dropout1			0
dense			6500
dropout			0
flatten			0
Saída			768802
<b>Total</b>			<b>795014</b>

Tabela 5 – Comparação das métricas de desempenho da FayNet vs FayNet2.0 com imagens FBIRN.

Redes	Métricas (%)				
	AC	PR	ESP	SEN	F2
FayNet	71,18 $\pm$ 6,69	60,75 $\pm$ 7,23	71,45 $\pm$ 6,36	71,12 $\pm$ 7,01	68,77 $\pm$ 7,12
FayNet 2.0	75,63 $\pm$ 4,28	75,42 $\pm$ 3,79	74,65 $\pm$ 2,93	77,88 $\pm$ 2,03	77,37 $\pm$ 2,91

Nota-se que para as mesmas condições descritas no trabalho de (VERGARA, 2018), não foram encontradas os mesmos resultados, obtivemos um resultado pouco melhor através do novo modelo, uma vez que as métricas da FayNet 2.0 estão melhores em comparação ao modelo original. Essa melhora pode ser vista em um cenário mais notório nas Figuras 22 e 23.

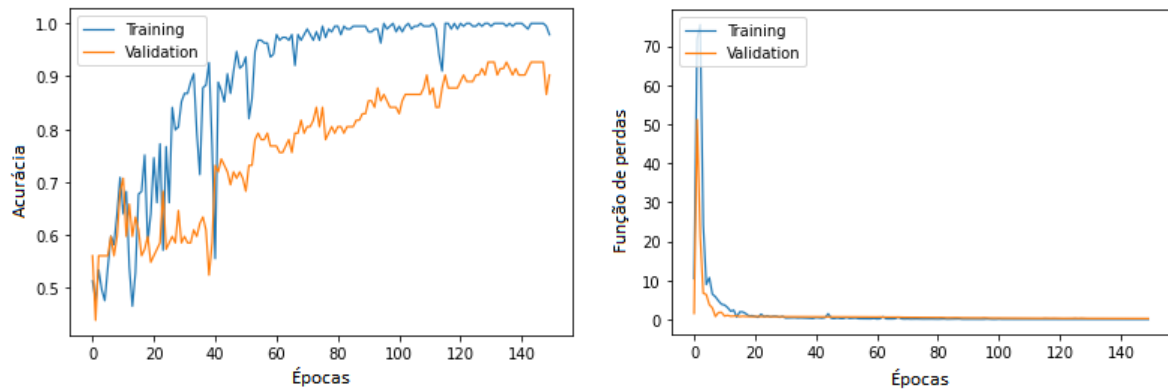


(a) Acurácia do treinamento e validação FayNet. (b) Perdas do treinamento e validação FayNet.

Figura 22 – Gráficos de desempenho da FayNet com o banco de dados FBIRN. Fonte: Autor.

Os gráficos representam a acurácia e as perdas durante o treinamento e validação dos modelos ao decorrer das épocas. Na Figura 22, é evidente uma maior oscilação no comportamento das curvas, tanto no treino como na validação, em relação a FayNet 2.0, conforme mostrado na Figura 23. Esse comportamento foi reduzido quando adicionado o método de *batch normalization*, que auxilia na estabilização durante a etapa de treinamento. No caso da troca da função de ativação da camada de *pooling* para a *Average Pooling*, faz com que a presença média dos recursos sejam destacados e não recursos específicos como no *pooling* máximo.

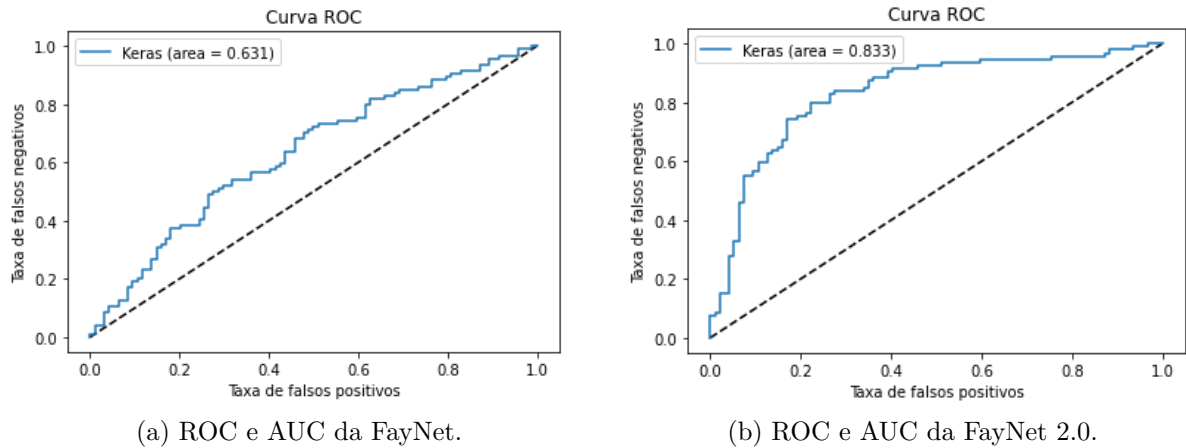
Uma observação é que algumas das imagens usadas neste teste, nos cortes especificados, não apresentavam a região de interesse, o que pode ter interferido nos resultados obtidos. Destaca-se também que alguns parâmetros não foram descritos pelo (VERGARA, 2018), como a função de perdas. Para ambos os modelos foi utilizada a função de entropia cruzada binária, função apropriada para a classificação binária. A FayNet 2.0 conseguiu um desempenho médio de 75,63% de acurácia com um valor de perda de 0,41, enquanto o modelo original adquiriu cerca de 71,18% de acurácia e 0,63 de perdas.



(a) Acurácia do treinamento e validação FayNet 2.0. (b) Perdas do treinamento e validação FayNet 2.0.

Figura 23 – Gráficos de desempenho da FayNet 2.0 com o banco de dados FBIRN. Fonte: Autor.

Outro método empregado para analisar o desempenho das redes foi a curva ROC, exibida na Figura 24. A curva ROC traça a probabilidade do sistema em classificar corretamente determinada classe. A área sob curva reflete a capacidade de classificação dos modelos, ou seja, a FayNet possui o potencial de 63,1% de classificar pacientes portadores de esquizofrenia e a FayNet 2.0 83,3%.



(a) ROC e AUC da FayNet.

(b) ROC e AUC da FayNet 2.0.

Figura 24 – Curva ROC dos modelos com o banco de dados FBIRN. Fonte: Autor.

Os resultados aqui estão claros quando observado o comportamento das curvas. Na Figura 24a está expressa a probabilidade da classificação da FayNet, onde pode-se observar que seu comportamento foge do comportamento característico ideal descrito na subseção 2.5.3. Esse achatamento visto no gráfico reflete um mal desempenho do classificador, uma vez que apresenta um pequeno distanciamento da linha central que representa uma classificação onde a taxa de verdadeiros positivos e falsos positivos atingem 50%. Já a Figura 24b expõe um comportamento de um classificador moderado, onde a orientação da curva possui uma maior distância da linha central quando comparado ao gráfico da Figura

Tabela 6 – Arquitetura da rede FayNet com os parâmetros para o banco de dados Schiz-Connect.

Nome (Camada)	Tamanho filtro/passo	Tamanho de saída	Parâmetros de aprendizado
Entrada		1X229x220	0
Conv2d	5X5/1	32X225X216	832
max_pooling2d	2X2/1	32X112X108	0
Conv2d	3X3/1	64X110X106	18496
max_pooling2d	2X2/1	64X55X53	0
dropout1			0
dense			6500
dropout			0
flatten			0
Saída			583002
<b>Total</b>			<b>608830</b>

Tabela 7 – Arquitetura da rede FayNet com os parâmetros para o banco de dados Schiz-Connect.

Nome (Camada)	Tamanho filtro/passo	Tamanho de saída	Parâmetros de aprendizado
Entrada		1X229X220	0
Conv2d	5X5/1	32X225X216	832
batch_normalization			128
average_pooling2d	2X2/1	32X112X108	0
Conv2d	3X3/1	64X55X53	18496
bach_normalization			256
averege_pooling2d	2X2/1	64X62X62	0
dropout1			0
dense			6500
dropout			0
flatten			0
Saída			583002
<b>Total</b>			<b>609214</b>

24a, o que confirma um melhor desempenho do novo modelo em relação a reconstrução do modelo original da FayNet.

Outro teste aplicado aos modelos foi a utilização dos bancos de dados disponibilizados pelo *SchizConnect* com o total de 1936 imagens, conforme descrito em 3.1. Neste teste também foram utilizadas as mesmas ferramentas de validação de performance no teste anterior, como o *K-fold*, métricas de desempenho e a curva ROC.

As particularidades destas arquiteturas estão refletidas nas Tabelas 6 e 7.

Após a realização do treinamento na etapa de teste com o novo conjunto de dados, foi possível obter uma acurácia de  $88,71 \pm 4,07\%$  com o modelo da FayNet2.0 e  $79,85 \pm$

2,29% com a FayNet original.

Nas Figuras 25 e 26 são traduzidas as curvas de validação e perda para cada um dos modelos. Vale notar que a FayNet 2.0 possui uma maior acurácia e menor perda tanto nas etapas de treinamento quanto nas etapas de validação.

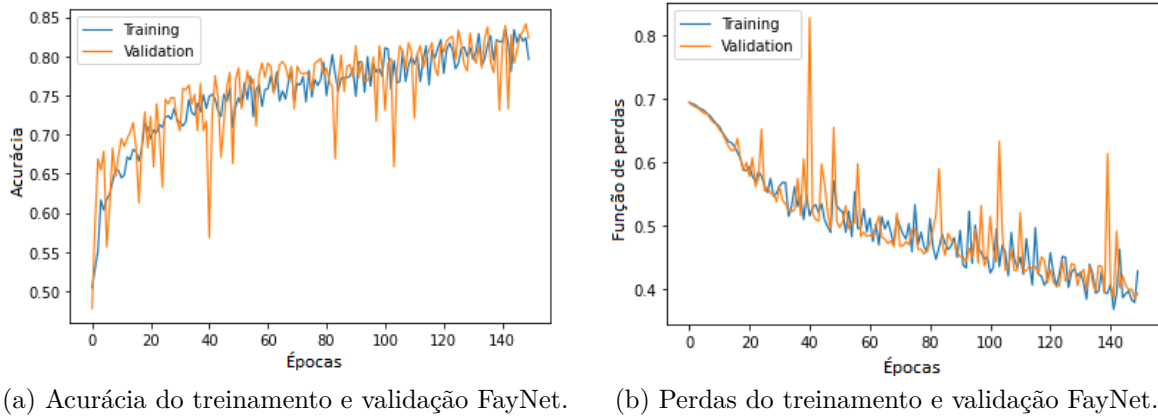


Figura 25 – Resultados do treinamento e validação da FayNet utilizando imagens disponibilizadas pelo SchizConnect. Fonte: Autor.

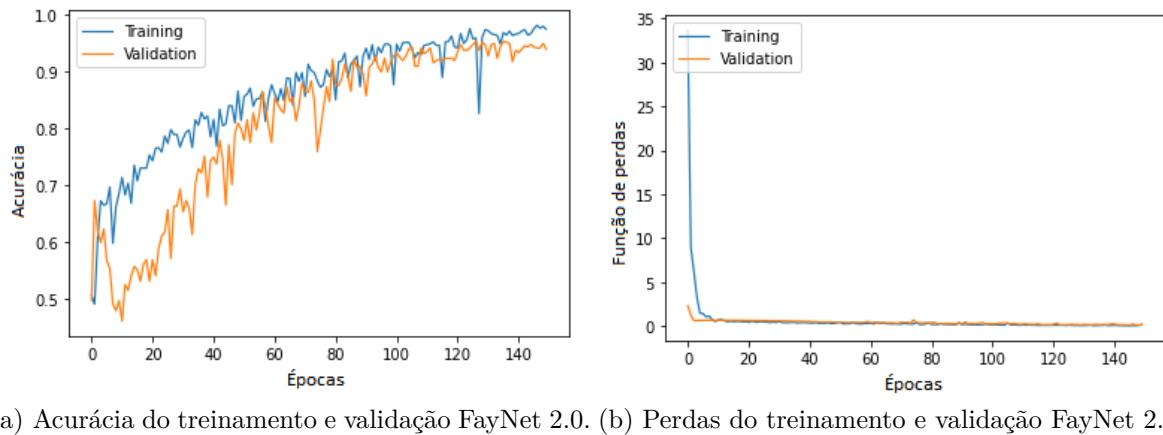


Figura 26 – Resultados do treinamento e validação FayNet 2.0 utilizando imagens disponibilizadas pelo SchizConnect. Fonte: Autor.

Ao comparar as curvas ROC entre as duas arquiteturas nas Figuras 24 e 27, além das métricas de desempenho nas Tabelas 5 e 8, nota-se a melhoria nos resultados ao aumentar a quantidade de dados fornecidos as redes, assim como o melhor desempenho da FayNet 2.0 em relação a FayNet original.

Um dado relevante é que as duas arquiteturas aqui foram treinadas com o mesmo otimizador, o SGD, e mesmo *batchsize*, 90 conforme explicado em 3.3.1. O *batchsize* foi determinado de forma empírica através de testes com o sistema. Percebeu-se que com um valor abaixo deste estabelecido, as perdas no conjunto são elevadas e em uma dada época,

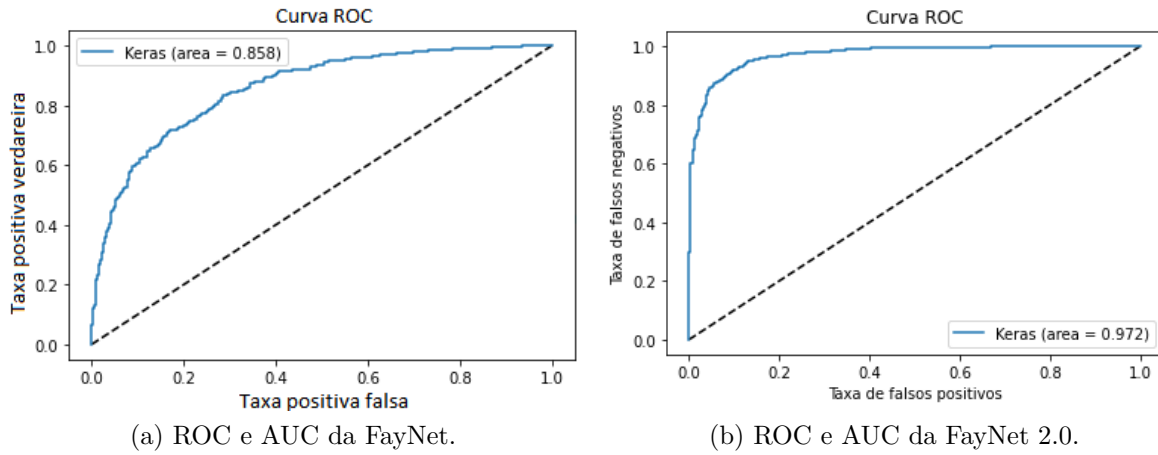


Figura 27 – Curva ROC dos modelos com o banco de dados SchizConnect. Fonte: Autor.

Tabela 8 – FayNet original vs FayNet2.0 utilizando imagens disponibilizadas pelo SchizConnect. Fonte: Autor.

Redes	Métricas (%)				
	AC	PR	ESP	SEN	F2
FayNet	79,95 $\pm$ 2, 29	79,48 $\pm$ 2, 49	79,52 $\pm$ 3, 07	80,25 $\pm$ 2, 32	80,09 $\pm$ 2, 4
FayNet 2.0	88,71 $\pm$ 4, 07	85,38 $\pm$ 3, 62	86,22 $\pm$ 3, 71	85,74 $\pm$ 3, 19	85,66 $\pm$ 3, 4

o comportamento da curva de perdas começa a crescer, indicando o *overfitting*. Já para valores acima de 90, a acurácia das redes decaía.

Dado o desempenho das redes nos testes submetidos, vale ressaltar que os resultados obtidos com o treinamento no banco de dados utilizado por Vergara (VERGARA, 2018), mesmo realizando a construção segundo o modelo proposto por ele, obtiveram uma acurácia inferior a obtida em seu trabalho. Isso pode se dar pelo fato de se utilizar um *framework* diferente, ou até mesmo devido ao não conhecimento de alguns hiper-parâmetros, como por exemplo, a função de perdas utilizada. Em ambos os modelos a função utilizada foi a *Binary crossentropy* por ser a função mais adequada para uma abordagem de classificação binária.

Logo, é importante apontar ainda que se os resultados obtidos pela FayNet neste trabalho fossem os mesmos reportados por Vergara, seria possível que o desempenho desta rede para o banco de dados *SchizConnect*, alcança-se um melhor desempenho, uma vez que a performance dos modelos aumentou quando adicionado mais dados.

## 4.3 Resnet

Com base nos resultados apresentados, é perceptível que as redes aqui testadas possuem grande potencial para realizar a classificação de imagens de MRI cerebral de pacientes portadores de esquizofrenia e auxiliar no diagnóstico dessa doença complexa.

Como no trabalho de Fung (FUNG et al., 2019), não apresentava os parâmetros da rede, o modelo foi testado empiricamente para tornar possível a verificação do desempenho de uma rede mais profunda para realização da tarefa de classificação, uma vez que o conjunto de imagens nos fornece parâmetros suficientes para justificar sua utilização.

Assim baseado no modelo da Figura 19, e nos testes empíricos realizados, a ResNet 18 apresentou um resultado médio de  $85,19 \pm 2,47\%$  de acurácia e  $30,8\%$  de perda, nos conjuntos de testes. A possibilidade de atualização do peso das camadas mais profundas, pode ser um dos fatores que auxiliam o bom desempenho dessa rede. Na Tabela 9, pode-se ainda observar algumas outras métricas para avaliação do modelo utilizado.

Dentre as métricas, a precisão é uma das formas mais fáceis de se avaliar um modelo. No entanto ela consegue apresentar a porcentagem das imagens classificadas corretamente e não leva em consideração algumas sutilezas, no que diz respeito aos custos diferentes de falsos negativos e falsos positivos.

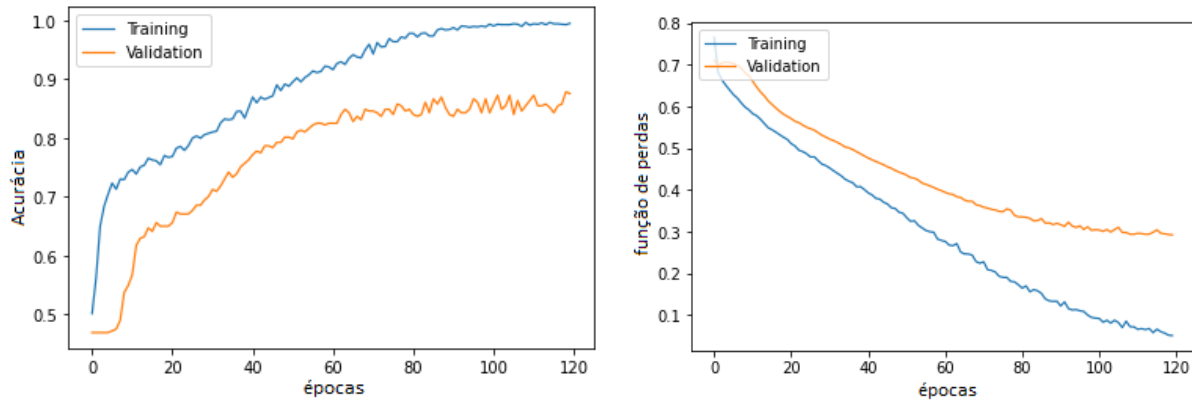
Assim para uma melhor avaliação do modelo a métrica *F2-score* foi calculada, uma vez que em análises médicas, seria pior não perceber um paciente com uma determinada doença do que dar positivo a um paciente saudável (ZHANG; DEMNER-FUSHMAN, 2017). Nesse contexto, o resultado da métrica *F2-score* foi de  $85,37 \pm 0,03\%$ , como mostrado na Tabela 9. Isso demonstra um bom equilíbrio do modelo entre a precisão e sensibilidade.

Tabela 9 – Métricas de desempenho para avaliação do modelo treinado da ResNet18.

Rede	Métricas (%)				
	AC	PR	ESP	SEN	F2
ResNet18	$85,19 \pm 2,47$	$85,52 \pm 0,02$	$85,18 \pm 0,02$	$85,34 \pm 0,03$	$85,37 \pm 0,03$

A partir da Figura 28a, pode-se observar que no conjunto de validação a rede apresentou uma saturação em torno de 84%, quando chega próximo a 80 épocas de treinamento. No entanto como a perda ainda apresentava uma queda significativa, assim 120 épocas foram utilizadas para tentar encontrar um mínimo de perdas possível.

Com o intuito de avaliar o modelo treinado, foi plotada a curva ROC e calculada a área sob essa curva, a AUC. O resultado está ilustrado na Figura 29, onde apresentou-se



(a) Acurácia do treinamento e validação ResNet18. (b) Perdas do treinamento e validação ResNet18.

Figura 28 – Resultados do treinamento e validação ResNet18 utilizando imagens disponibilizadas pelo SchizConnect. Fonte: Autor.

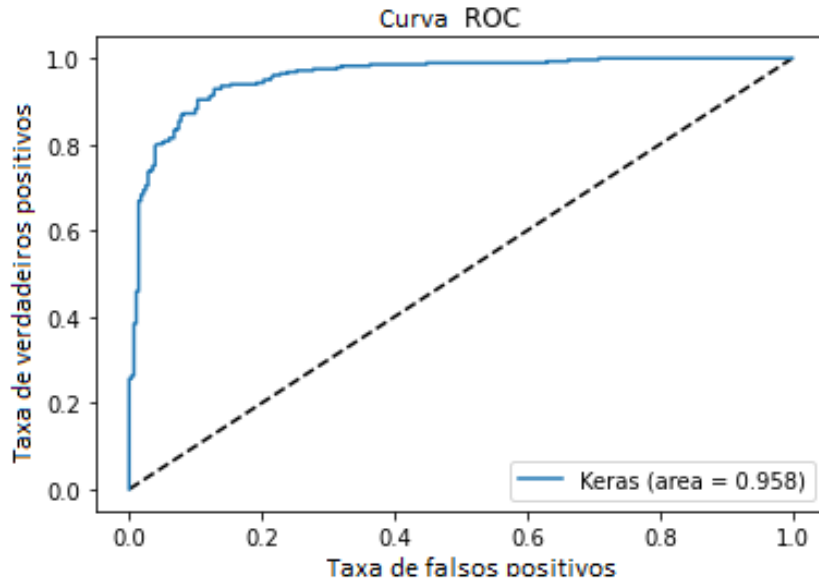


Figura 29 – ROC e AUC referente ao teste do modelo da ResNet18. Fonte: Autor.

uma AUC igual a 0,958 e uma curva ROC que apresenta uma distância considerável da linha de 50% de probabilidade, mostrando que o modelo de classificação possui uma boa taxa de verdadeiros positivos e baixa taxa de falsos positivos.

Comparando-se os resultados obtidos com o uso do banco de dados *SchizConnect*, pode-se analisar também o resultado de cada classificador. Na Tabela 10 é abordado o valor médio das métricas de desempenho obtidas por cada arquitetura, bem como o respectivo desvio padrão.

Nota-se que o modelo FayNet 2.0 apresentou os melhores resultados, como uma



Tabela 10 – Tabela de comparação dos modelos descritos. Fonte: Autor.

Redes	Métricas (%)				
	AC	PR	ESP	SEN	F2
FayNet	79,95 $\pm$ 2, 29	79,48 $\pm$ 2, 49	79,52 $\pm$ 3, 07	80,25 $\pm$ 2, 32	80,09 $\pm$ 2, 4
FayNet 2.0	88,71 $\pm$ 4, 07	85,38 $\pm$ 3, 62	86,22 $\pm$ 3, 71	85,74 $\pm$ 3, 19	85,66 $\pm$ 3, 4
ResNet18	85,19 $\pm$ 2, 47	85,52 $\pm$ 0, 02	85,18 $\pm$ 0, 02	85,34 $\pm$ 0, 03	85,37 $\pm$ 0, 03

acurácia média de classificação de 88,71 % e especificidade de 86,22 %, o que traduz o modelo a uma boa análise à classe dos pacientes saudáveis.



## 5 Conclusão

Este trabalho propôs uma análise comparativa de redes neurais convolucionais para classificação de esquizofrenia. Os modelos FayNet e ResNet18 foram implementadas e testadas com o banco de dados público *SchizConnect*, a partir das descrições de sua arquitetura. Porém, como alguns parâmetros não estavam disponíveis, a reconstrução não foi exatamente fiel. No caso da CNN FayNet, criada por Vergara (VERGARA, 2018), observou-se um comportamento não satisfatório encontrado nos seus resultados, dando a possibilidade da realização de algumas alterações em seus parâmetros, como o inclusão do método de *batch normalization*, que apresentaram melhores resultados tanto durante o treinamento quanto na realização dos testes. Os resultados encontrados mostraram que a classificação de esquizofrenia com as CNNs apresentadas são bastante significativos, como uma elevada média de acurácia em comparação á resultados de classificação de tarefas semelhantes relatados nos trabalhos citados.

O novo modelo, a FayNet 2.0, obteve uma acurácia de classificação média de 88,71% e a ResNet18 85,19%. Um dos fatores essenciais que auxiliaram na obtenção desses resultados foi a etapa de pré-processamento como o *software* Clínica. Nele alinhou-se as imagens para que a região cerebral de interesse fosse a mesma, visto que o banco de dados é composto pela integração de várias instituições de aquisição de MRI. Para maior confiabilidade dos modelos, foi utilizada o método de avaliação cruzada *k-fold* para a avaliação do desempenho das redes, com 7 separações de dados de treinamento e testes diferentes.

Assim, com base nos resultados obtidos a partir da utilização das CNN's na classificação de MRI cerebral, distinguindo pacientes portadores de esquizofrenia de saudáveis, foi possível notar o potencial dessas arquiteturas utilizadas, e a viabilidade da continuação de estudos, com o intuito de aumentar o desempenho dessas redes, seja melhorando suas arquiteturas ou pré-processamentos utilizados.

Como trabalhos futuros, pode-se sugerir:

- Testar a nova arquitetura com o pré-processamento de segmentação das imagens, onde o cérebro é separado do crânio.
- Aumentar o número de classes do modelo, para a rede classificar as diferentes variações de esquizofrenia, dada pelo banco de dados *SchizConnect*.
- Utilizar a nova rede para a classificação de doenças mentais com alterações morfológicas semelhantes, como por exemplo o transtorno Bipolar.



# Referências

- ABADI, M. et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. 2015. Software available from tensorflow.org. Disponível em: <<https://www.tensorflow.org/>>. Citado 2 vezes nas páginas 55 e 56.
- ABRAHAM, A. et al. Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, v. 8, p. 14, 2014. ISSN 1662-5196. Disponível em: <<https://www.frontiersin.org/article/10.3389/fninf.2014.00014>>. Citado na página 54.
- ALFONSE, M.; SALEM, A.-B. M. An automatic classification of brain tumors through mri using support vector machine. *Egy. Comp. Sci. J*, v. 40, n. 3, 2016. Citado na página 30.
- ALPAYDIN, E. *Machine learning: the new AI*. [S.l.]: MIT press, 2016. Citado na página 31.
- ANDREASEN, N. C. et al. Progressive brain change in schizophrenia: a prospective longitudinal study of first-episode schizophrenia. *Biological psychiatry*, Elsevier, v. 70, n. 7, p. 672–679, 2011. Citado na página 29.
- Antioquia, A. M. C. et al. Zipnet: Zfnet-level accuracy with 48× fewer parameters. In: *2018 IEEE Visual Communications and Image Processing (VCIP)*. [S.l.: s.n.], 2018. p. 1–4. Citado 2 vezes nas páginas 41 e 42.
- ASSOCIATION, A. P. *Diagnostic and Statistical Manual of Mental Disorders*. 5. ed. [S.l.]: American Psychiatric Publishing, 2013. Citado na página 24.
- ASSOCIATION, A. P. et al. *DSM-5: Manual diagnóstico e estatístico de transtornos mentais*. [S.l.]: Artmed Editora, 2014. Citado 3 vezes nas páginas 24, 27 e 28.
- AVANTS, B. B. et al. The insight toolkit image registration framework. *Frontiers in Neuroinformatics*, v. 8, p. 44, 2014. ISSN 1662-5196. Disponível em: <<https://www.frontiersin.org/article/10.3389/fninf.2014.00044>>. Citado na página 54.
- BERRAR, D. Cross-validation. In: \_\_\_\_\_. [S.l.: s.n.], 2018. ISBN 9780128096338. Citado na página 46.
- BISHOP, C. M. *Pattern Recognition and Machine Learning*. 5. ed. [S.l.]: Springer Science, 2006. ISBN 0-387-31073-8. Citado na página 31.
- BORA, M. B. et al. Handwritten character recognition from images using cnn-ecoc. In: . [S.l.: s.n.], 2020. v. 167, p. 2403–2409. Citado na página 40.
- Budhiman, A.; Suyanto, S.; Arifianto, A. Melanoma cancer classification using resnet with data augmentation. In: *2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*. [S.l.: s.n.], 2019. p. 17–20. Citado 2 vezes nas páginas 13 e 43.
- CAMPINAS, C. R. *Ressonância magnética*. 2015. Disponível em: <<http://www.crcamp.com.br/index.php/exames/ressonancia-magnetica>>. Citado na página 29.

- CHAKRABORTY, D. et al. Assessment and prediction of negative symptoms of schizophrenia from rgb+ d movement signals. In: IEEE. *2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)*. [S.l.], 2017. p. 1–6. Citado na página 24.
- Cheng Lv, J.; YI, Z.; TAN, K. Global convergence of oja's pca learning algorithm with a non-zero-approaching adaptive learning rate. *Theoretical Computer Science*, v. 367, n. 3, p. 286–307, 2006. ISSN 0304-3975. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S030439750600421X>>. Citado na página 46.
- CIRESAN, D. C. et al. Flexible, high performance convolutional neural networks for image classification. In: *Twenty-second international joint conference on artificial intelligence*. [S.l.: s.n.], 2011. Citado na página 34.
- CLEMENTZ, B. A. et al. Identification of distinct psychosis biotypes using brain-based biomarkers. 2016. Citado 2 vezes nas páginas 27 e 28.
- CONSORTIUM, G. de Trabalho de Esquizofrenia do P. G. Insights biológicos de 108 loci genéticos associados à esquizofrenia. 2015. Citado na página 27.
- DAS, S. *CNN Architectures: LeNet, AlexNet, VGG, GoogLeNet, ResNet and more...*. 2017. Disponível em: <<https://medium.com/analytics-vidhya/cnns-architectures-lenet-alexnet-vgg-googlenet-resnet-and-more-666091488df5>>. Citado 2 vezes nas páginas 13 e 42.
- DEMILY, C.; FRANCK, N. Cognitive remediation: a promising tool for the treatment of schizophrenia. *Expert Review of Neurotherapeutics*, Taylor & Francis, v. 8, n. 7, p. 1029–1036, 2008. Citado na página 23.
- DEVARRIYA, D. et al. Unbalanced breast cancer data classification using novel fitness functions in genetic programming. *Expert Systems with Applications*, Elsevier, v. 140, p. 112866, 2020. Citado na página 48.
- DHOMNE, A.; KUMAR, R.; BHAN, V. Gender recognition through face using deep learning. *Procedia computer Science*, v. 132, p. 2–10, 2018. Citado 3 vezes nas páginas 13, 42 e 43.
- DIELEMAN, S. et al. Lasagne: First release. *Aug*, 2015. Citado na página 55.
- EFFENDY, N. et al. Artificial neural network prediction on ultrasonic performance of bismuth-tellurite glass compositions. *Journal of Materials Research and Technology*, v. 9, n. 6, p. 14082 – 14092, 2020. ISSN 2238-7854. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S2238785420318330>>. Citado na página 32.
- ERP, T. G. van et al. Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the enigma consortium. *Molecular psychiatry*, Nature Publishing Group, v. 21, n. 4, p. 547–553, 2016. Citado na página 29.
- FERRARI, T. B. Volumetria cerebral de pacientes esquizofrênicos. 2008. Citado na página 29.

- FONOV, V. et al. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, v. 47, p. S102, 2009. ISSN 1053-8119. Organization for Human Brain Mapping 2009 Annual Meeting. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1053811909708845>>. Citado na página 54.
- FUNG, Y. R. et al. Alzheimer's disease brain mri classification: Challenges and insights. *arXiv preprint arXiv:1906.04231*, 2019. Citado 3 vezes nas páginas 54, 56 e 69.
- GLOVER, G. H. et al. Function biomedical informatics research network recommendations for prospective multicenter functional mri studies. *Journal of Magnetic Resonance Imaging*, v. 36, n. 1, p. 39–54, 2012. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/jmri.23572>>. Citado na página 55.
- GOMES, G. S. d. S.; LUDERMIR, T. B. Redes neurais artificiais com funções de ativação complemento log-log e probit para aproximar funções na presença de observações extremas. *Revista da Sociedade Brasileira de Redes Neurais (SBRN)*, v. 6, n. 2, p. 142–153, 2008. Citado na página 38.
- GONZALEZ, R. C.; WOODS, R. E.; EDDINS, S. L. *Digital image processing using MATLAB*. [S.l.]: Pearson Education India, 2004. Citado 2 vezes nas páginas 30 e 31.
- GOODFELLOW, I. et al. *Deep learning*. [S.l.]: MIT press Cambridge, 2016. v. 1. Citado 3 vezes nas páginas 32, 34 e 37.
- GORGOLEWSKI, K. J. et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific data*, Nature Publishing Group, v. 3, n. 1, p. 1–9, 2016. Citado na página 53.
- GUA, J. et al. Recent advances in convolutional neural networks. *ROSE Lab, Interdisciplinary Graduate School, Nanyang Technological University, Singapore*, 2017. Citado na página 40.
- GUYON, I. et al. Unipen project of on-line data exchange and recognizer benchmarks. 1994. Citado na página 41.
- HAN, C. et al. Learning smooth representations with generalized softmax for unsupervised domain adaptation. *Information Sciences*, v. 544, p. 415 – 426, 2020. ISSN 0020-0255. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0020025520308458>>. Citado na página 38.
- HAYKIN, S. *Neural Networks: A Comprehensive Foundation (3rd Edition)*. USA: Prentice-Hall, Inc., 2007. ISBN 0131471392. Citado na página 33.
- He, K. et al. Deep residual learning for image recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2016. p. 770–778. Citado na página 43.
- HINTON, G. E.; SHALLICE, T. Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological review*, American Psychological Association, v. 98, n. 1, p. 74, 1991. Citado na página 33.
- HINTON, G. E. et al. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*, 2012. Citado na página 39.

- HUBEL, D. H.; WIESEL, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, Wiley-Blackwell, v. 160, n. 1, p. 106, 1962. Citado na página 34.
- IMAGING, N. I. of B.; NIBIB, B. *Magnetic Resonance Imaging (MRI)*. 2018. Disponível em: <<https://www.nibib.nih.gov/science-education/science-topics/magnetic-resonance-imaging-mri>>. Citado na página 28.
- IOFFE, S.; SZEGEDY, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *CoRR*, abs/1502.03167, 2015. Disponível em: <<http://arxiv.org/abs/1502.03167>>. Citado na página 37.
- JIANG, R. *Implementing ResNet-18 Using Keras*. [S.l.: s.n.], 2020. <<https://www.kaggle.com/songrise/implementing-resnet-18-using-keras>>. Citado na página 56.
- KARPATHY, A. Stanford university cs231n: Convolutional neural networks for visual recognition. URL: <http://cs231n.stanford.edu/syllabus.html>, 2018. Citado 4 vezes nas páginas 13, 35, 37 e 38.
- KHAN, S. et al. *A Guide to Convolutional Neural Networks for Computer Vision*. [S.l.]: Morgan Claypool Publishers, 2018. ISBN 9781681730226. Citado 2 vezes nas páginas 32 e 33.
- KHODAEI, A.; HOSSEIN-ZADEH, G.-A.; ANANLOO, E. S. Comparison of volumes of subcortical regions in schizophrenia patients and healthy controls using mri. In: IEEE. *2015 2nd International Conference on Pattern Recognition and Image Analysis (IPRIA)*. [S.l.], 2015. p. 1–5. Citado 3 vezes nas páginas 23, 28 e 29.
- KIRCHER, T. et al. Formal thought disorders: from phenomenology to neurobiology. *The Lancet Psychiatry*, Elsevier, v. 5, n. 6, p. 515–526, 2018. Citado na página 24.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep. 2012. Citado 3 vezes nas páginas 13, 40 e 41.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, ACM New York, NY, USA, v. 60, n. 6, p. 84–90, 2017. Citado 2 vezes nas páginas 13 e 37.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. *Commun. ACM*, Association for Computing Machinery, New York, NY, USA, v. 60, n. 6, p. 84–90, maio 2017. ISSN 0001-0782. Disponível em: <<https://doi.org/10.1145/3065386>>. Citado na página 41.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, v. 521, p. 436–444, 2015. Citado na página 34.
- LECUN, Y. et al. Backpropagation applied to handwritten zip code recognition. *Neural computation*, MIT Press, v. 1, n. 4, p. 541–551, 1989. Citado na página 34.
- Lecun, Y. et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, v. 86, n. 11, p. 2278–2324, 1998. Citado 2 vezes nas páginas 13 e 40.



- LECUN, Y.; K., K.; C., F. Convolutional networks and applications in vision. in circuits and systems (iscas). *Proceedings of 2010 IEEE International Symposium on (2010)*, v. 591, p. 253–256, 2010. Citado na página 40.
- LECUN, Y.; KAVUKCUOGLU, K.; FARABET, C. Convolutional networks and applications in vision. In: IEEE. *Proceedings of 2010 IEEE international symposium on circuits and systems*. [S.l.], 2010. p. 253–256. Citado 2 vezes nas páginas 35 e 36.
- LESHNO, M. et al. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks*, v. 6, n. 6, p. 861 – 867, 1993. ISSN 0893-6080. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0893608005801315>>. Citado na página 34.
- LIN, T.-Y. et al. Focal loss for dense object detection. In: \_\_\_\_\_. [S.l.: s.n.], 2018. Citado na página 45.
- MARTINS, V. E. Aplicação de deep learning para detecção de veias em carne suína. 2018. Citado na página 44.
- MATHUR, Y.; JAIN, P.; SINGH, U. Foremost section study and kernel support vector machine through brain images classifier. In: IEEE. *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)*. [S.l.], 2017. v. 2, p. 559–562. Citado na página 30.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. *Journal of Materials Research and Technology*, v. 5, n. 4, p. 115–133, 1943. ISSN 1522-9602. Citado na página 33.
- NAIR, V.; HINTON, G. E. Rectified linear units improve restricted boltzmann machines. In: *ICML*. [S.l.: s.n.], 2010. Citado na página 36.
- NIFTI. *NifTI: — Neuroimaging Informatics Technology Initiative*. 2020. Disponível em: <<https://nifti.nimh.nih.gov/>>. Citado na página 51.
- NOGUEIRA, L. D. Segmentação de imagem de ressonância magnética de cérebro baseada em slic superpixel. 2019. Citado 2 vezes nas páginas 24 e 29.
- NOURI, D. n. scikit-learn compatible neural network library. 2014. Citado na página 55.
- OLIVEIRA, R. M.; FACINA, P. C. B. R.; JÚNIOR, A. C. S. La realidad del vivir con esquizofrenia. *Revista Brasileira de Enfermagem*, SciELO Brasil, v. 65, n. 2, p. 309–316, 2012. Citado na página 23.
- Reverdy, P.; Leonard, N. E. Parameter estimation in softmax decision-making models with linear objective functions. *IEEE Transactions on Automation Science and Engineering*, v. 13, n. 1, p. 54–67, 2016. Citado na página 38.
- ROMAGUERA, L. V. et al. Segmentação do miocárdio em imagens de mri cardíaca utilizando redes neurais convolutivas. Universidade Federal do Amazonas, 2017. Citado na página 39.
- SAIKIA, A. R. et al. Comparative assessment of cnn architectures for classification of breast fnac images. In: . [S.l.: s.n.], 2019. v. 57, p. 8–14. Citado na página 40.

- SAYED, A. M.; ZAGHLOUL, E.; NASSEF, T. M. Automatic classification of breast tumors using features extracted from magnetic resonance images. *Procedia Computer Science*, Elsevier, v. 95, p. 392–398, 2016. Citado na página 30.
- SHARMA, A.; RAMKIRAN, S. Mri based schizophrenia patient classification: a meta-cognitive approach. In: IEEE. *2015 International Conference on Cognitive Computing and Information Processing (CCIP)*. [S.l.], 2015. p. 1–6. Citado na página 24.
- SILVA, A. M. et al. Esquizofrenia: Uma revisão bibliográfica. *UNILUS Ensino e Pesquisa*, v. 13, n. 30, p. 18–25, 2016. Citado 2 vezes nas páginas 23 e 24.
- SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. Citado na página 42.
- SMITH, R. C. et al. Nuclear magnetic resonance in schizophrenia: a preliminary study. *Psychiatry Research*, Elsevier, v. 12, n. 2, p. 137–147, 1984. Citado na página 29.
- SOUZA, L. A. d.; COUTINHO, E. S. F. Fatores associados à qualidade de vida de pacientes com esquizofrenia. *Brazilian Journal of Psychiatry*, SciELO Brasil, v. 28, n. 1, p. 50–58, 2006. Citado na página 23.
- SRIVASTAVA, N. et al. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, JMLR. org, v. 15, n. 1, p. 1929–1958, 2014. Citado 2 vezes nas páginas 13 e 39.
- SURIANI, J. et al. Offline signature verification using deep learning convolutional neural network (cnn) architectures googlenet inception-v1 and inception-v3. 2019. Citado na página 42.
- TEAM, T. D. Theano: A python framework for fast computation of mathematical expressions. arxiv e-prints abs/1605.02688. 2016. Citado na página 55.
- TORRES, U. d. S. *Estudo de alterações estruturais cerebrais em pacientes com esquizofrenia crônica e de primeiro episódio através de imagens por ressonância magnética com morfometria baseada no voxel*. Tese (Doutorado) — Universidade de São Paulo, 2016. Citado 2 vezes nas páginas 28 e 29.
- UNIVERSITY, D. T. Convolutional neural networks (cnns / convnets). 2019. Citado na página 42.
- VERGARA, R. F. Detecção de alterações cerebrais anatômicas associadas à esquizofrenia com base em redes convolucionais aplicadas a imagens de ressonância magnética. 2018. Citado 14 vezes nas páginas 13, 24, 29, 32, 35, 39, 44, 54, 55, 56, 62, 64, 68 e 73.
- WANG, L. et al. *Schizconnect*. 2018. Disponível em: <<http://schizconnect.org/>>. Citado 4 vezes nas páginas 13, 15, 51 e 52.
- Appendix a - theoretical foundations. In: WITTEN, I. H. et al. (Ed.). *Data Mining (Fourth Edition)*. Fourth edition. Morgan Kaufmann, 2017. p. 533–552. ISBN 978-0-12-804291-5. Disponível em: <<https://www.sciencedirect.com/science/article/pii/B9780128042915000234>>. Citado na página 45.

- YADAV, S.; SHUKLA, S. Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification. In: *2016 IEEE 6th International Conference on Advanced Computing (IACC)*. [S.l.: s.n.], 2016. p. 78–83. Citado na página 46.
- YANG, Z. et al. Vegetable image retrieval with fine-tuning vgg model and image hash. 2018. Citado na página 42.
- ZHANG, K.; DEMNER-FUSHMAN, D. Automated classification of eligibility criteria in clinical trials to facilitate patient-trial matching for specific patient populations. *Journal of the American Medical Informatics Association*, Oxford University Press, v. 24, n. 4, p. 781–787, 2017. Citado na página 69.
- ZHANG, Y. et al. Plaintext-related image encryption algorithm based on perceptron-like network. *Information Sciences*, v. 526, p. 180 – 202, 2020. ISSN 0020-0255. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0020025520302346>>. Citado 2 vezes nas páginas 13 e 33.
- Zhang, Z. et al. Cnn-fl: An effective approach for localizing faults using convolutional neural networks. In: *2019 IEEE 26th International Conference on Software Analysis, Evolution and Reengineering (SANER)*. [S.l.: s.n.], 2019. p. 445–455. Citado na página 34.
- ZHANG, Z. et al. A study of effectiveness of deep learning in locating real faults. *Information and Software Technology*, p. 106486, 2020. ISSN 0950-5849. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0950584920302287>>. Citado na página 34.
- ZHOU, Y.-T. et al. Image restoration using a neural network. *IEEE transactions on acoustics, speech, and signal processing*, IEEE, v. 36, n. 7, p. 1141–1151, 1988. Citado na página 37.